

Centuries of Knowledge
Graduate School of Library and Information Science
Data Curation Education Program

IMLS RE-05-05-0036

Final Report
October 1, 2006 – December 31, 2011

Allen Renear, Ph.D.
Professor and Interim Dean
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
DCEP PI (August 16-December 30, 2011)

Melissa Cragin, Ph.D.
AAAS Science & Technology Policy Fellow
National Science Foundation
Directorate for Biological Sciences
(Formerly Research Assistant Professor at Illinois)
DCEP PI (December 1, 2009-August 15, 2011)

P. Bryan Heidorn, Ph.D.
Director
School of Information Resources and Library Science
University of Arizona
(Formerly Associate Professor at Illinois)
DCEP PI (October 1, 2006-November 30, 2009)

Carole L. Palmer, Ph.D.
Professor and Director
Center for Informatics Research in Science & Scholarship
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
DCEP Co-PI

Linda C. Smith, Ph.D.
Professor and Associate Dean for Academic Programs
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
DCEP Co-PI

John Unsworth, Ph.D.
Vice Provost for Library & Technology Services
Chief Information Officer
Brandeis University
(Formerly Dean of the Graduate School of
Library and Information Science at Illinois)
DCEP Co-PI

Megan Senseney, MSLIS
Project Coordinator, Research Services
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
DCEP Project Coordinator

Abstract

The Centuries of Knowledge grant was designed to increase educational and research capacity in data curation at the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. We developed the Data Curation Education Program, a specialization within our Master of Science degree program, graduating 38 students to date. New courses developed for the specialization include Foundations of Data Curation, a survey course on the emerging field, and Digital Preservation. We developed the Summer Institute on Data Curation for practicing information professionals, facilitating the development of a community of practice across U.S. and Canadian academic and research organizations. Our outreach and service activities have led to a range of new partnerships that have resulted in student fieldwork opportunities, as well as new collaborative research and education activities resulting in 4 successful grant proposals.

Table of Contents:

Narrative Report

1. Introduction.....	3
2. Curriculum Development.....	3
3. Recruiting and Student Financial Support.....	5
4. Building Community Capacity.....	5
5. Outreach and Training.....	7
6. Local, National, and International Leadership.....	8
7. Analysis, Lessons Learned, Accomplishments.....	9

Appendices

Appendix A: Advisory Committee Members.....	12
Appendix B: DCEP Core Courses and Recommended Electives.....	13
Appendix C: Course Syllabi and Case Study.....	15
1. LIS590DC Foundations of Data Curation (Spring 2012).....	15
2. LIS586 Digital Preservation (Spring 2012).....	31
3. Case Study in Data Curation: Crystallography.....	41
Appendix D: Recruitment Materials.....	44
Appendix E: List of Disseminations.....	52
Appendix F: Summer Institute Programs.....	57
1. 2008 Summer Institute on Data Curation.....	57
2. 2010 Summer Institute on Data Curation.....	59
3. 2011 Summer Institute on Data Curation.....	61

1. Introduction

The primary goal of the Centuries of Knowledge grant was to design a program of graduate study to serve as a model for training data curators within the context of a Library and Information Science (LIS) education. In support of that goal, the project team also integrated ongoing research and practice into graduate training to foster the development of information specialists who understand the research culture and can make substantive contributions to the mission of scientific, humanities, social science, and cultural heritage institutions and libraries.

The specific objectives were to:

- 1) Develop curriculum for the specialization that builds on existing graduate programs at GSLIS;
- 2) Establish internships and develop a job network at institutions where students can develop and apply their growing expertise;
- 3) Expand understanding of the role of data curation in the production of research; and
- 4) Share the educational approach with other schools interested in developing similar specializations.

In August 2008, our first student to complete the Specialization in Data Curation graduated with an MSLIS. As of December 2011, 38 students have completed the specialization and graduated, another 10 students have completed the specialization and are expected to graduate in May 2012, and 60 currently enrolled students have expressed interest in the program and are pursuing data curation coursework.

Results of grant-supported activities include:

- 1) A master's specialization with a set of new and updated courses covering issues in data curation;
- 2) A Summer Institute short-course for continuing education of practicing information professionals;
- 3) Field-based practical experience for data curation students;
- 4) Integration of course work with current informatics research at GSLIS, other UIUC departments, and collaborating institutions; and
- 5) Dissemination of curriculum materials, philosophy, and design through conference presentations and service activities.

2. Curriculum Development

We have developed a program to train a new generation of LIS professionals for positions with responsibility for handling and stewardship of digital research data. The specialization in Data Curation was developed over several semesters through an iterative review and update process by the grant team, based on internal assessment, feedback from the Advisory Committee, and results of informal consultations with internal and external colleagues. During the program development process we increased the required "Core Courses" and modified the set of "Recommended Electives". The specialization is offered as an option within our Master of Science degree program, and all required courses are available via GSLIS's on-line learning option, LEEP. This has allowed students from across the country to complete the specialization.

Appendix B provides course descriptions for all DCEP Core Courses and Recommended Electives.

We offered the specialization beginning in 2007, following the development of two new required “core” courses. The **Foundations of Data Curation** (LIS590DCL) is a survey course that provides students with a balanced view of the emergent curation landscape. **Digital Preservation** (LIS586 formerly, LIS590PD) is a conceptual course, addressing current problems and approaches to preservation systems and services. Both of these courses have generated great interest across the GSLIS student body, and by year three of the grant, each course was being offered in both the fall and spring semesters, with the latter being enrolled at maximum capacity on a regular basis. Syllabi for each of the above-mentioned courses are included in Appendix C along with a data curation case study conducted while developing the DCEP curriculum.

Three new courses were developed to support the specialization that are open to all students in the school as well as students in our campus-wide Bioinformatics master’s degree program: Information Transfer and Collaboration in Science (LIS590TR), Ontology Development (LIS590OD), and an ontologies seminar entitled Ontologies in the Natural Sciences (LIS590ON).¹ Finally, several established GSLIS courses were modified to include a greater emphasis on data curation and to align curatorial perspectives and knowledge with the primary theme of the class. Introduction to Biological Informatics Problems and Resources (LIS530BI, formerly LIS590BI) is one example of a course that was re-designed to contribute to the data curation curriculum and students who have interests in biology. This course also provides greater exposure to data curation concepts and practices for students in the Bioinformatics M.S. program. Additional courses modified to address data curation-related concepts include Systems Analysis and Design (LIS453), which is a required “Core Course” for the specialization, and Information Modeling (LIS590IM), which is a “Recommended Elective”.

Pre-professional Development

Current LIS pedagogy and input from our Advisory Committee were the basis for developing a pool of field placement partners to support practical training experiences. We utilized two formal mechanisms for fieldwork: practica and internships. The practicum functions as a for-credit course that requires 100 hours of hands-on experience on-site in a place of the student’s choosing; these generally serve as an introduction to an organization, and result in a small project or paper. The internship is a full-time placement in an organization, and generally includes a project of a larger scope. Students do not receive course credit, and the DCEP provided a stipend and living expenses during these placements.

During the first year of the grant project, we identified four sites with internship and practicum opportunities for students interested in data curation: the Smithsonian Institution, the National Library of Medicine, Purdue University Library, and Johns Hopkins Library; in subsequent years, Woods Hole Oceanographic Institute and the National Snow and Ice Data Center (NSIDC) were also identified. From summer 2008 to summer 2011, we placed students in seven internships: one at each of the above-mentioned institutions, except NSIDC, which has hosted three interns. Faculty supervised seven practica between summer 2009 and summer 2011.

¹ Ontologies in Humanities has also been offered, and additional information will be available through the reports from the DCEP-H grant, IMLS RE-05-08-0062-08.

3. Recruiting and Student Financial Support Pre-professional Development

Recruitment

In collaboration with the GSLIS Communications staff, we developed a flyer and pamphlet about DCEP. These were distributed during general recruiting activities held at GSLIS, at professional events, and during school-wide recruiting trips. In addition, these materials were posted in several academic departments across our campus, and distributed at conferences we attended. We also compiled mailing lists, and on multiple occasions sent PR and recruiting materials to:

- a) McNair Scholars Program sites in selected colleges and universities throughout the United States;
- b) U.S. colleges and universities with undergraduate library and information science programs; and
- c) Biology and related departments at Historically Black Colleges and Universities.

Within GSLIS, DCEP faculty and staff made presentations about the program at the Orientation Day held at the start of each new academic year. Staff also visited class sessions in courses with high potential for recruits. One of our most successful in-house recruiting efforts is the on-going DCEP Meet and Greet, held annually during our fall on-campus “LEEP weekend,” required for distance education students. This event allows both LEEP and on-campus students interested in pursuing the Specialization in Data Curation to meet with current data curation students and faculty. Attendance at this event is always high, and now draws around 30 students, including recent graduates who come to talk about their experiences in the specialization. Samples of recruiting materials are included in Appendix D.

Student Financial Support

We have provided financial support for 60 students using several approaches, some of which were identified as we developed a better understanding of programmatic constraints and the need to have flexible options. Scholarships in the form of stipends were awarded to 34 students based entirely on financial need. Fourteen students were supported with fellowships, providing a stipend and tuition waiver. These were generally allocated through a competitive application process, but also designated based on merit to support students with exceptional professional promise. DCEP funds were also used to support travel for 12 students to data curation activities when the students were involved in the program or presenting a paper or poster.

4. Building Community Capacity

GSLIS Faculty Development

DCEP funding provided course release for two GSLIS faculty to develop curriculum: John MacMullen (2009) and Carole Palmer (2011). Professor MacMullen worked on creating stronger links between the data curation and life science informatics areas of the GSLIS curriculum. His activities included modifying the syllabus and content of LIS530BI “Introduction to Biological Information Problems and Resources,” and contributing to planning and developing the data curation curriculum. Professor Palmer worked on several activities to develop additional capacity for student education and training, including facilitating the relationship with a new data analytics specialization, revising the curriculum for LIS590DCL Foundations of Data Curation, advising two independent studies with doctoral students preparing to teach in the field of data

curation, coordinating curation and data science education efforts, and contributing to the campus-wide Illinois Data Stewardship Committee.

Needs Assessment

Over the span of the project, we conducted a series of formal and informal assessment activities to develop our knowledge base on researchers' data management practices, skills, and curation service needs. Investigations addressed a range of topics including data handling during the research lifecycle, storage and archiving, data sharing and ownership, and skills for data professionals.

Data Management Needs for UIUC

During the first year of the project, we partnered with a faculty member from the Department of Natural Resources and Environmental Sciences (NRES) to develop a Needs Assessment in collaboration with the (now disbanded) Environmental Council (EC) at the University of Illinois at Urbana-Champaign. The goals of this initial study were two-fold: The EC wanted to understand how data resources were being used across UIUC to inform requirements for data storage and support services; and, we were interested in the nature and extent of associated data management practices and problems to inform DCEP curriculum and learn about professional workforce needs.

A questionnaire was developed to survey the Faculty of the Environment, a self-identified cross-campus group with interests in environmental concerns, coordinated through the EC. Once the assessment protocol was approved by the Institutional Review Board, DCEP staff conducted a pre-test of the survey with a sample from the study population; the survey was modified based on the pre-test feedback, and then sent out to the Faculty of the Environment list. The survey results were augmented by a few additional faculty members who provided more in-depth information on the value and use of their data collections. Qualitative results were the most informative and were incorporated into ongoing research on differences in sub-disciplinary data practices. The findings helped shape our presentations on the philosophical base for the DC curriculum, and were also reported in conference posters and panel presentations.

Job Descriptions Assessment

When we began to collect job descriptions for curation-related positions in 2007, we found very few jobs that might be considered appropriate for LIS students. Five years later, we have seen a considerable change, concomitant with a broadening of curation activities in library and research organizations. We began an analysis of science-oriented job advertisements during the summer of 2009, and presented our initial findings in a poster at the International Digital Curation Conference in December of that year. We also collaborated with DCEP-H, a related IMLS grant project designed to extend the Data Curation Education Program to the humanities, to do a comparative analysis with a similar set of job ads data for the humanities. These data are now being used by researchers in the Center for Informatics Research in Science and Scholarship (CIRSS) to design a GSLIS-wide jobs analysis initiative that is being organized by Catherine Blake, Associate Director of CIRSS.

Professional Skills in Data Centers

The final phase of the Needs Assessment was a focused study of the skills and practices of data scientists and other data professionals at a national science data center. We conducted 13 interviews and have finished most of the coding, but this work was put on hold due to staff changes on the project. We had originally planned to complete a paper reporting the results in 2011, but Melissa Cragin now intends to continue this study beyond the conclusion of the grant.

5. Outreach and Training

The faculty and staff of DCEP participated in 24 data curation-related conferences and workshops over the span of the grant, providing 31 presentations through papers, panels, posters, and invited talks on the GSLIS data curation program and our related research activities. Participation at the annual International Digital Curation Conference (IDCC) and the American Society for Information Science and Technology (ASIS&T) conference was particularly strong, with the DCEP team often presenting at several sessions per meeting. In addition, we were invited to give talks about data curation by a variety of organizations beyond LIS, including the publishing sector and at scientific meetings, greatly extending awareness of our curriculum and potential roles for academic libraries in the handling of research data. A comprehensive list of conference and service activities is provided in Appendix E.

Dissemination about the curriculum was a significant activity, and each presentation contained information about the philosophical underpinnings of the specialization, the orientation to the emerging DC field, and an overview of the core courses and fieldwork opportunities. We responded to information requests from external faculty by supplying syllabi to UCLA and Rutgers and providing consultation on program start-up and course development to the University of Alberta.

Data Curation Summer Institutes (2008, 2010, 2011²)

Starting in the fall of 2007, practicing librarians began making requests to DCEP faculty and staff for training or classes in data curation. To address this need, we developed the Summer Institute for Data Curation, a short-course for LIS professionals who were actively preparing to or currently working with research data in their university library. This Summer Institute series has now run for four years. We continue to host a listserv for Summer Institute participants, which is currently moderated by Nic Weber, a GSLIS doctoral student and research assistant with the IMLS-funded Data Curation Education in Research Centers (DCERC) program.

The first institute was held June 2-5, 2008, with a focus on the general topic of scientific data and developing data curation services in the academic library. There were 30 participants and 8 presenters at this inaugural institute. Guest speakers presented a broad range of topics, including metadata and standards, preservation, intellectual property, appraisal and selection, and service planning.

The second DCEP-supported institute was held May 17-20, 2010, with a focus on geosciences. GSLIS hosted 24 practicing professionals from library and research organizations from across

² A Summer Institute was held in 2009 that was facilitated by Dr. Allen Renear and funded by the DCEP-H project; information on that workshop can be found in public reports for IMLS grant RE-05-08-0062-08.

the country and Canada, and DCEP provided travel support and reduced registration fees for 12 attendees. Topics for this Institute included principles and best practices for data curation activities; current metadata and format standards for the Earth and environmental sciences; data publishing; technical aspects of data repository systems and repository development planning; and approaches to partnering with domain scientists and data communities.

The third DCEP-supported institute was held June 6-9, 2011, with a focus on the life sciences. Although the funding for this event came primarily from the National Science Foundation through the Data Conservancy project, DCEP was integral to the planning and execution of the program, and Melissa Cragin led the four-day workshop. There were 24 attendees and 9 presenters from across the U.S. and one from Canada. Programs from each of the three institutes are included in Appendix F.

6. Local, National, and International Leadership

As part of the Chancellor's "Stewarding Excellence @ Illinois" initiative, Melissa Cragin served on the "IT at Illinois" Working Group. Part of the final report of this effort was a section on the need for a campus-wide approach to the handling of research data, developed in part to raise awareness of the coming mandates from federal funding agencies. One outcome of the Chancellor's initiative was the formation of the Illinois Research Data Initiative and the subsequent "Year of Data Stewardship" at the University of Illinois at Urbana-Champaign.

In the fall of 2010, Carole Palmer, Melissa Cragin, and Bryan Heidorn planned and led a national workshop, the "Data Curation Workforce Summit."³ While the Data Conservancy sponsored this meeting, DCEP supported Professor Heidorn's participation, and the presenters and audience were drawn, in part, from the network we had developed during the growth of the DCEP. The program focused on research data workforce needs from the government perspective, and participants included representatives from government agencies involved in data management and data curation and educators who are active in digital curation, together with representatives from NSF DataNet Projects actively involved in education programs for data curation. There were two objectives for the Summit: to build awareness and synergies among the educational activities and to provide exposure to important perspectives from potential employers.

IDEA

In December 2008, following the 4th Annual International Digital Curation Conference, Bryan Heidorn, Carole Palmer, and Allen Renear attended the International Data Curation Education Working Group Meeting, which was led by a team from the School of Information and Library Science at UNC. The IDEA workshop provided an opportunity for educators from many countries to compare educational methods and materials, and Palmer presented on DCEP progress. There was a notable degree of variability in training methods related to differences in target audience, differences in materials behind the digital surrogates, and differences in institution. The workshop provided a forum where data curation educators could reach an initial agreement on sharing information and course materials. The workshop also provided a forum for establishing collaborations for later projects. Allen Renear attended the follow-on IDEA meeting at the DigCCurr Conference in Chapel Hill, NC, on April 1, 2009.

³ The final report is available through the IDEALS repository at <http://hdl.handle.net/2142/28355>.

International Digital Curation Conference

In fall 2010, Melissa Cragin and Allen Renear (Project PIs for DCEP and DCEP-H) served as Program Co-chairs, along with Liz Lyon of UKOLN and Kevin Ashley of the DCC, for the Sixth International Digital Curation Conference (IDCC), which was co-hosted by GSLIS in Chicago, December 6-8, 2010. The program included 10 research papers, 9 practice papers, 35 posters and 3 demonstrations. It was very well received, and with 277 registered participants attendance far exceeded that of any previous year. GSLIS students had very strong participation at the conference. Aaron Collie, GSLIS DCEP graduate and Digital Curation Librarian at Michigan State University Libraries, led a team of GSLIS Data Curation students to produce an “amplified conference”, which he had begun coordinating and planning with the Digital Curation Centre while he was a GSLIS student. This group produced multimedia content that served as the online presence for the conference, including interviews; live feeds of presentations, photos, blog and Twitter entries; and production of archival video for the Digital Curation Centre.

7. Analysis, Lessons Learned, and Accomplishments

The DCEP project’s contribution to developing data curation capacity in the LIS community has been substantial and extensive. We have succeeded in our original goal of designing a program of graduate study at GSLIS that could serve as a model for training data curators within the context of LIS education. There is great demand for our students in this emerging area, and our graduates are being hired into new positions across a range of organizations, from academic libraries to public broadcasting agencies. Moreover, the project has left an impact on the LIS community that reaches well beyond our initial objectives. For example, we have put forward a definition of “data curation” that has helped to stimulate the examination of the role of librarians in the research enterprise. Evidence of this impact can be seen both in and outside of our field, from LIS blog posts⁴ to business publications⁵.

Advisory Committee

From February 2007 to November 2008, we convened 4 meetings with the members of our advisory committee. The goal of these meetings was to establish what LIS students needed to learn to become professional data curators and to develop case studies and a set of best practices for developing data curation expertise. The committee consisted of experts from a range of different disciplinary domains and was instrumental in early development and capacity building for faculty and staff, which occurred during the first three years of the grant. Afterward, we continued to maintain interactions with specific board members while general committee activity tapered. In retrospect, consulting the advisory committee was less effective during the execution phase when the DCEP team worked on planning outreach activities within target organizations, recruiting, and developing curricular materials. A complete list of advisory committee members is included in Appendix A.

⁴ Dempsey, L. (2007, February 25). Data curation education. Retrieved from <http://orweblog.oclc.org/archives/001277.html>.

⁵ Hemsoth, N. (2012, January 12). The evolving art (and business) of data curation. *datanami*. Retrieved from http://www.datanami.com/datanami/2012-01-12/the_evolution_of_data_curation.html.

Formalization of the Program

From the start of the grant, the DCEP program used a team approach to design the requirements for a specialization in data curation. The team included several GSLIS faculty, the Project Coordinator, and Graduate Research Assistants who met regularly to discuss the formulation of course requirements and curriculum. Over the course of this iterative process, the necessary technical foundation for professional work in data curation was articulated, and the team identified a set of Recommended Electives (of which a minimum of two are required to complete the specialization) to address these technical needs.

Throughout the 2011-2012 academic year, we have continued assessing the Specialization in Data Curation in the context of the general GSLIS curriculum, and we are re-evaluating the best way to formalize the program within the School. The team holds monthly meetings to ensure a smooth transition to full institutional program support now that the DCEP project has drawn to a close. Our primary concerns include maintaining student recruitment and advising activities, and formalizing administrative aspects of the specialization. We have, for example, created a student declaration form to be submitted by students who intend to complete the specialization, and now provide a letter of acknowledgement for all students who have completed the program upon graduation. Please refer to Appendices K and L for samples of the form and letters. Moving forward, we are considering offering the Specialization in Data Curation as a GSLIS-based certificate program. We had previously considered a formal transcribed concentration, managed by the University's Graduate College, but a locally managed certificate allows the flexibility for periodic program updates needed for curriculum in an emerging field.

Recruitment

When we wrote the grant proposal, we identified two programs at GSLIS that would provide a unique opportunity to recruit underserved communities to the LIS profession: the undergraduate minor in information technology studies and our award-winning online masters program, LEEP. As anticipated, our undergraduate program has proven to be an excellent introduction to LIS for a range of students from different backgrounds. Through LEEP, students can attend from anywhere in the world who are unable to leave their homes and travel to central Illinois due to financial, work, or family constraints. Since the LEEP program requires visits to the University of Illinois campus once per semester, we made an important revision to the budget to include support for travel and lodging for students from underserved groups with economic need who are coming to campus as part of the LEEP program.

During the first two years of the project, we focused recruitment activities on students from underserved populations by contacting programs at outside institutions. We also targeted science departments on our own campus. Our recruitment packets included a letter, flyer, and poster sent to each department or program. The flyer provided an overview of the DC concentration, and included required coursework, scheduling options, application procedures and financial aid information. The poster highlighted the program in more general ways, by inviting readers to think about what will be required for digital information to remain accessible in the future. While we cannot attribute any specific applications to our mailings, DCEP was often indicated as a program of interest by people considering GSLIS, and we now have high-quality applicants who come to GSLIS specifically for the data curation program. Toward the end of the grant, our main recruitment efforts have focused on new and currently enrolled GSLIS students; and in our first

year of targeted internal recruitment overall interest in the data curation program increased by 100%.

Developing Field Placement Opportunities

Field placements provide LIS students with the opportunity to apply theoretical concepts to real-world problems in an organizational setting. Although the Specialization in Data Curation was developed within the Master of Science degree at GSLIS, the program has also attracted students from GSLIS pursuing the post-masters Certificate of Advanced Study and students pursuing the Master of Science in Bioinformatics. These students often possess a level of technical expertise that makes them a great match for the internship opportunities that we developed as part of the GSLIS portfolio of field placement opportunities. Their early participation in internships and practica proved invaluable as it helped us establish relationships with external institutions while we were still developing the data curation curriculum. Our relationship with NSIDC has proven especially successful. In the past year, NSIDC hired two of our student interns for full-time positions; Elizabeth Schlagel was hired immediately upon graduation and Lynn Yarmey was recruited from a related position in an academic library.

Throughout the grant period, project staff played an active role in identifying field placement opportunities, establishing relationships with outside institutions, and allocating resources to support to students seeking placement. Moving forward, we will be shifting responsibility for data curation field placement opportunities from DCEP staff to the primary point-of-service offices within GSLIS.

Outreach and Training

DCEP outreach and training activities have had a broad impact, promoting professional conversations about the role and level of engagement that academic and research libraries ought to have in the research process, and for curation and stewardship of research data in particular. Fostering these ideas has proven significant both for the individuals attending these events and for the growth of the field. Capacity has been increased across many organizations for taking on curatorial activities and stewardship of what are (for most libraries) a new type of collection.

Impacts of the Summer Institute include, for example:⁶

- Launch of a working group on organizational data service needs and capacity, in which the leader “relied heavily on what I had learned at the Summer Institute for orienting the group members and proposing objectives.”
- Creation of a series of lunch time workshops on writing data management plans for NSF.
- Development of an initiative to capture data associated with master’s theses and PhD dissertations that involves working with the institutional repository, the Graduate Division, and the Vice Chancellor for Research and Graduate Education.
- Collaboration with faculty and library administration on a campus-wide survey to understand data needs of faculty and inform library service development for data access and use.

⁶ These are drawn from an informal survey of Summer Institute attendees.

Appendix A: Advisory Committee Members

Chris Freeland, Application Development Manager and Project Manager, Missouri Botanical Garden (Active 2006-2010)

Thomas Garnett, Associate Director for Digital Library and Information Systems, Smithsonian Institution (Active 2006-2010)

Gen. William D. Goran, U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory (ERDC-CERL) (Active 2006-2007)

Martin Kalfatovic, Head of the New Media Office and Preservation Services Department, Smithsonian Institution Libraries (Active 2006-2010)

Maryann Martone, Ph.D., Co-Director for the National Center for Microscopy and Imaging Research (NCMIR), University of California, San Diego (Active 2006-2007)

Joanna McCaffrey, Collections Database Architect, The Field Museum (Active 2006-2010)

Nancy McGovern, Digital Preservation Office, Inter-University Consortium for Political and Social Research (ICPSR), University of Michigan (Active 2007-2010)

Katherine McNeill-Harman, Data Services and Economics Librarian, Massachusetts Institute of Technology (Active 2006-2010)

Chuck Miller, Vice President, Information Technology and Chief Information Officer, Missouri Botanical Garden (Active 2006-2010)

Indra Neil Sarkar, Ph.D., Informatics Manager, Marine Biological Laboratory (Active 2006-2007)

Chris Rewerts, Ph.D., U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory (ERDC-CERL) (Active 2006-2007)

David Soller Ph.D., Geologist, U.S. Geological Survey (Active 2007-2010)

Appendix B: DCEP Core Courses and Recommended Electives

Core Courses

LIS590DCL Foundations of Data Curation*
Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities and policies enable data discovery and retrieval, maintain data quality and add value, and provide for re-use over time. This course provides an overview of a broad range of theoretical and practical problems in this emerging field. Examines issues related to appraisal and selection, long-lived data collections, research lifecycles, workflows, metadata, legal and intellectual property issues.
LIS586 Digital Preservation*
Examines current problems with and approaches to digital preservation that are fundamental to the long-term accessibility of digital materials. Also examines the range of current research problems, along with emerging methods and tools, and assess a variety of organizational scenarios to plan and implement a preservation plan. Topics will include basic information theory, preservation of complex digital objects; standards and specifications; sustainability and risk assessment; authenticity, integrity, quality control, and certification; and management of preservation activities.
LIS453 Systems Analysis and Management
Covers how to evaluate, select and manage the information systems that will be used in the daily operation of libraries and information centers. Includes the systems used by technical staff and the information consumers. Course will focus on information as a product. Attention is given to the operation of an organization as a whole and the impact of change on the integration of resources, work flow and usability. Formal methods for modeling systems, and industry practice techniques of analysis are used to address these problems and opportunities.

Recommended Electives

LIS452 Foundations of Information Processing in Library and Information Science
Covers the common data and document processing constructs and programming concepts used in library and information science. The history, strengths and weaknesses of the techniques are evaluated in the context of our discipline. These constructs and techniques form the basis of applications in areas such as bibliographic records management, full text management and multimedia. No prior programming background is assumed.
LIS490DB Introduction to Databases
The course provides students with both theoretical and practical training in good database design. By the end of the course students will create a conceptual data model using entity-relationship diagrams, understand the importance of referential integrity and how to enforce data integrity constraints when creating a database. Students will be proficient in writing basic queries in the structured query language (SQL) and have a general understanding of relational database theory including normalization.

LIS590DI Digital Libraries: Research and Practice
A comprehensive examination of the history and state-of-the-art in digital library research and practice. Focuses upon the theoretical, technological, human factors and evaluative components of digital library research and practice. Course includes an intensive reading of the literature, review of existing technologies and proof-of-concepts implementation projects. This course is foundational for students wishing to engage seriously in the world of digital librarianship. Students should have access to a personal computer upon which they can experiment on their own with downloaded software tools. Students must be competent in basic computing including the installation and configuration of software packages.
LIS590IM Information Modeling
An introduction to the principles of information modeling commonly used to support digital library applications such as collections management and electronic publishing. The course takes a logic-based approach to analyzing and comparing different modeling methods. Specific modeling practices covered include relational database design, entity relationship modeling (ER/EER), document grammars (XML), and semantic web languages (RDF/S and OWL).
LIS590MD Metadata in Theory and Practice
Metadata plays an increasingly critical role in the creation, distribution, management and use of electronic materials. This course will combine theoretical examination of the design of metadata schema with their practical application in a variety of settings. Hands-on experience in the creation of descriptive, administrative and structural metadata, along with their application in systems such as OAI harvesting, OpenURL resolution systems, metasearch systems and digital repositories, will help students develop a thorough understanding of current metadata standards as well as such issues as crosswalking metadata schema, metadata's use in information retrieval and data management applications, and the role of standards bodies in metadata schema development.
LIS590OD Ontology Development
An introduction to formal ontology focusing on development and implementation issues and contemporary ontology software tools and languages. In spring of 2008 we will use as example ontologies one for museum and heritage information (CIDOC-CRM) and one for biological information (the Functional Model of Anatomy). Students may also do projects on other ontologies in other areas if they wish. The ontology editor Protege will be used throughout and the representation of ontologies in W3C semantic web languages RDF(S) and OWL will be emphasized.
LIS590RO Representing and Organizing Information Resources
Emphasizes concepts and methods of organizing information resources across different settings and systems, or within one particular setting. The course extends the basic conceptual foundation provided in LIS 501 by providing further reading, analysis, discussion, and practice related to one or several major traditions of information organization in different environments (e.g., libraries, museums, archives, Internet, and within a single organization).

* Syllabus included in Appendix C.

Appendix C: Course Syllabi and Case Study**Foundations of Data Curation**

LIS590DCL- Spring 2012

Synchronous sessions:

Thursdays 4-6pm (Central time)

Instructors:*Carole Palmer* (clpalmer@illinois.edu)

Office hours: Thursdays 6-7pm or by appointment

Office: LISB 314; Phone: 217.244.0653;

Skype: clpalm400

Allen Renear (renear@illinois.edu)

Office hours: Thursdays 6-7pm or by appointment

Office: LISB 303; Phone: 217.265.5216

Teaching assistants:*Tiffany Chao* (tchao@illinois.edu)Office hours: Mondays 4-5pm (via skype: tiffcchao)
or by appointment*Simone Sacchi* (sacchi1@illinois.edu)Office hours: Tuesdays 4-5pm (via skype chat:
simone.sacchi) or by appointment**Semester (at-a-glance)**

Date	Week	Topic	Assignments
19 Jan	Week 1	Introduction	
26 Jan	Week 2	Data Curation Landscape	
2 Feb	Week 3	Context, Concepts, and Models	
9 Feb	Week 4	Data Practices of Researchers (guest speaker: Suzie Allard, <i>University of Tennessee-Knoxville</i>)	
16 Feb	Week 5	Building Data Collections	Assignment 1 due
23 Feb	Week 6	Representation Levels	
3 March	Week 7	(LEEP on-campus) Interpretation Levels	
8 March	Week 8	Models and Systems	
15 March	Week 9	Guest speaker: <i>Ruth Duerr</i> on Earth Science (<i>National Snow and Ice Data Center</i>)	Assignment 2 due
22 March		Spring Break	
29 March	Week 10	Data Management Planning (guest speaker: Sarah Shreeves, <i>University of Illinois Libraries</i>)	
5 April	Week 11	Policy and Regulatory Environment	
12 April	Week 12	Infrastructure (guest speaker: Mark Evans, <i>Tessella</i>)	Assignment 3 due
19 April	Week 13	Guest speaker: <i>Julia Flanders</i> on Humanities (<i>Brown University Women Writers Project</i>)	
26 April	Week 14	(Final Class)	
7 May		Monday @ noon	Assignment 4 due

Course content subject to change. Please refer to the course Moodle page for updates.

Participation (20%)

Participation is an important part of the course and of being a good colleague and citizen of the class.

Class attendance: Attendance is required for synchronous live sessions and the on-campus LEEP day. Absence of a synchronous session should be reported to instructors as soon as possible.

Class participation: Students are expected to complete readings and assignments for the week prior to the class session and to actively engage and contribute to synchronous class sessions in a way that demonstrates that preparation.

Forum posts and discussion: There are several opportunities to discuss and share your thoughts on the Moodle site. Most of your grade for participation will be assessed based on your Reading Comments and Responses. It is important to make early and consistent contributions and to have a voice on the discussion forums. Do not hold off until later in the semester to contribute your comments and responses.

- *Reading Comments:* Provide constructive, substantive comments for a minimum of 3 weeks. For each week, one of the assigned readings has been selected (designated by *) as the focus of discussion. Comments do not need to be exclusively on the selected paper, but we encourage deeper engagement among the class as a whole around this item.
 - Comments should NOT be a summary of what you have read. They should contribute to thoughtful analysis of the content and its applications, and to dialogue among the group. These postings will usually consist of a paragraph or two, and might explain a particularly interesting or important “take away” or new insight for you, raise questions or difficulties you had with a concept or argument, or discuss relationships to previous class material or forum comments. Discussion is perpetual. There are no hard deadlines for when reading comments are due for each week, since discussion during class and postings may stimulate further thinking and commenting.
- *Responses:* You are expected to offer constructive responses to the Reading Comments offered by your colleagues, and, in doing so, to engage with them in a respectful manner. Responses should be more frequent than Reading Comments, and reflect your growing knowledge of the course content.
- *General discussion:* You are encouraged to post and discuss news items, events, articles, or other information related to data curation to the general discussion forum, to help build continual awareness of research and practice.

Assignment 1: Comparison of Digital Data Collections (15%)

[DUE: Thursday, February 16th by 11:55pm]

The aim of this assignment is to become familiar with research data collections as information resources for research communities and objects of curation. The exercise applies the functional categories for digital data collections—Research, Resource, and Reference—from the 2005 National Science Board report, *Long-Lived Digital Data Collections*, an assigned reading for Week 2 of the course.

Select two collections for comparison.

Briefly describe each collection based on available information, and provide a URL. Identify which of the three categories you think applies best to the collection.

Compare the two collections on the following features: audience, scope, standards, curation, maintenance, funding, and persistence. For many collections this information may not be explicit and may need to be inferred from what you can determine from its origin, content, and any available documentation or literature.

Discuss and compare the collections considering the following questions:

- How well do the collections fit into the category you assigned? Why were the other categories not appropriate?
- What challenges did you encounter determining the category? What other characteristics might be important for understanding curation needs and long-term use by researchers? How do these characteristics compare for the two collections?
- What are the implications for curation services for the categories? How much variation might we see within a category across collections?

The following are potential starting points for locating individual, named collections:

Scientific Data Repositories, compiled by Laura Marcial and Brad Hemminger, School of Information & Library Science, UNC-Chapel Hill, spring 2010, available:
http://www.lib.unc.edu/reference/data_services/researchdatatoolkit/repositories.html#social_sciences [excel spreadsheet]

Arts & Humanities Data Services: <http://www.ahds.ac.uk/>

Social Science Data Archives: <http://www.sociosite.net/databases.php>

Deliverable: Your analysis should be approximately 400-600 words and may extend beyond the guiding questions as space allows. Tables or figures may be used, but include explanations of the information presented in them. Include references to any sources or documentation reviewed. Please use APA citation style.

Submit under 'Assignments' on the Moodle as Word (.docx). Please use the following file naming convention: [last name]_A[assignment #]_DCL12.docx (e.g. Sacchi_A1_DCL12.docx)

Grading rubric: evidence of in-depth examination of collections and supporting documentation; evidence of understanding of categories; quality of analysis; completeness of response to guiding questions.

Assignment 2- Representation Levels (20%)

[DUE: Thursday, March 15th by 11:55pm]

The objective of this assignment is to strengthen your understanding of the topics discussed in the class sessions titled Representation Levels, Interpretation Levels, and Models & Systems.

- 1) Select either a documented data vocabulary or a documented "data format" that is used to support scholarly work in the sciences or humanities. You may choose a metadata vocabulary if you wish, as long as it is not Dublin Core. Do not choose an image format; although you may choose an image-oriented metadata vocabulary.
- 2) Find at least one concrete example of this data vocabulary or data format being used to present or describe real data.
- 3) Very briefly (2-4 paragraphs) characterize your chosen data vocabulary or format, and your example. Provide references and urls to all relevant documentation. Also, if possible: indicate what appear to be the 1-5 most authoritative articles that discuss the nature of your vocabulary or format, and 1-5 of the most important articles or that analyze problems or propose improvements or changes.
- 4) Discuss (2-4 paragraphs) your selected vocabulary or format. Summarize and abstract, do not document individual features except as examples needed to indicate the general nature.
- 5) Discuss (2-4 paragraphs) your selected example of your vocabulary or format.

In 4) and 5) be sure to,

- a) Give particular attention to the four representation levels in the "Simple Stack": (i) Semantics, (ii) Syntax, (iii) Serialization, (iv) Character Encoding.
- b) Include some discussion of possible alternative representations of the same information,
- c) Include some discussion of data curation issues such as preservation and re-use.

NB: Your discussions will vary substantially in nature depending on how much of the Simple Stack is specified in the vocabulary or format you chose. If you chose an abstract metadata vocabulary (e.g. Dublin Core) or an ontology such CIDOC-CRM then the separation of semantics and serialization will probably be clearly and explicitly anticipated in the standard, and little may be said in the documentation itself about possible serialization strategies -- although in your example some serialization strategy or other will have been chosen. On the other hand if you chose a data format such as CIF (Crystallographic Information File), TEI (Text Encoding Initiative), or any of the XML gene or protein file formats then semantics and serialization will be specified in the documentation, but the degree with which the distinction (semantics vs. serialization) is recognized in the documentation will vary and alternative serializations may not have been anticipated. That should not necessarily be an obstacle to identifying, describing, and discussing those alternatives.

Submit under 'Assignments' on the Moodle as Word (.docx). Please use APA citation style. Please use the following file naming convention: [last name]_A[assignment #]_DCL12.docx (e.g. Sacchi_A2_DCL12.docx)

Assignment 3: Data Management Plans (20%)

[DUE: Thursday, April 12th by 11:55pm]

The aim of this assignment is to learn about the range of resources available to support data management planning and to get experience evaluating and applying these resources.

Scenario: You are part of a newly formed data management consulting team. As a data curation professional, it is your responsibility to evaluate and provide recommendations on data management plans developed by researchers and research groups for their proposal submissions to funding agencies.

Funding agencies have begun to provide “guidelines” on their expectations for data management plans, and research libraries and data centers are developing “resources” as a service to assist researchers with the development of data management plans and data management more generally.

- Select a sample data management plan from the list below.
- Identify a data management resource suitable for guiding evaluation of the plan.
- Evaluate the plan based on the selected resource. Be sure to account for the guidelines set by the appropriate funding agency.
- Assess one other resource as an alternative resource for guiding the evaluation process.

Deliverables:

1) Report addressed to the researcher who wrote the plan providing feedback (such as strengths and weaknesses of the plan presented) and recommendations on their plan, based on one of the resources and agency guidelines. It may also be helpful to pose questions for the researcher to consider as part of the comments. This document should be written from the perspective of an information professional communicating with a member of their service community.

2) A memo to your data management team documenting the similarities and difference of the two resources and their usefulness for evaluating data management plans. This document should be written from the perspective of an information professional communicating with their colleagues.

Responses for each deliverable should be approximately 500-600 words. Please use APA citation style.

You may use a data management plan or set of guidelines not listed below, but it must be approved by Instructor Chao in advance. Submit the responses under ‘Assignments’ on the Moodle as a Word (.docx).

Please use the following file naming convention: [last name]_A[assignment #]_DCL12.docx (e.g. Sacchi_A3_DCL12.docx)

Grading rubric: Evidence of review and comprehension of selected plan and resources; appropriate application of guidelines and resources; appropriate, helpful, and clearly stated feedback and recommendations; sound analysis and comparison of guidelines.

Sample data management plans

DM Plan for depositing data with ICPSR:

[HTTP://WWW.ICPSR.UMICH.EDU/ICPSRWEB/CONTENT/ICPSR/DMP/PLAN.HTML](http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/DMP/PLAN.HTML)

Natural Science Data Management plan examples:

[HTTP://WWW.ICPSR.UMICH.EDU/ICPSRWEB/ICPSR/DMP/RESOURCES.JSP#Ao6](http://www.icpsr.umich.edu/icpsrweb/ICPSR/DMP/RESOURCES.JSP#Ao6)

U-Wisconsin: [HTTP://RESEARCHDATA.WISC.EDU/MAKE-A-PLAN/EXAMPLES/](http://researchdata.wisc.edu/make-a-plan/examples/)

Rice University (plans for Social and Behavior Science, Biosciences, and Physical Sciences and Engineering):

[HTTP://OSR.RICE.EDU/FORMS/DATAMANAGEMENTPLANS.PDF](http://osr.rice.edu/forms/datamanagementplans.pdf)

DataONE sample plans. Available: [HTTP://WWW.DATAONE.ORG/PLANS](http://www.dataone.org/plans)

Data management planning guidelines

All the institutions listed (national and international) at Association of Research Libraries (ARL) Data Management Planning. (2011). [HTTP://WWW.ARL.ORG/RTL/ERESearch/ESCIEIN/NSF/NSFRESOURCES.SHTML](http://www.arl.org/rtl/eresearch/escien/NSF/NSFRESOURCES.SHTML)

MIT "Data Planning Checklist". [HTTP://LIBRARIES.MIT.EDU/GUIDES/SUBJECTS/DATA-MANAGEMENT/CHECKLIST.HTML](http://libraries.mit.edu/guides/subjects/data-management/checklist.html)

JHU Data Management Questionnaire. [HTTP://DMP.DATA.JHU.EDU/JHUDATAMANAGEMENT](http://dmp.data.jhu.edu/jhudatamanagement)

DataONE. [HTTP://WWW.DATAONE.ORG/PLANS](http://www.dataone.org/plans)

Australian National Data Service: data management for Researchers.

[HTTP://ANDS.ORG.AU/RESEARCHERS/MANAGE-DATA.HTML](http://ands.org.au/researchers/manage-data.html)

Digital Curation Centre: Data Management Plans. [HTTP://WWW.DCC.AC.UK/RESOURCES/DATA-MANAGEMENT-PLANS](http://www.dcc.ac.uk/resources/data-management-plans)

CIESIN: Geospatial Electronic Records. [HTTP://WWW.CIESIN.COLUMBIA.EDU/GER](http://www.ciesin.columbia.edu/ger)

Oak Ridge National Laboratory: Best Practices for Preparing Environmental Data Sets to Share and Archive. [HTTP://DAAC.ORNL.GOV/PI/BESTPRAC.HTML](http://daac.ornl.gov/pi/bestprac.html)

Borer, E.T., et al. (2009). Some Simple Guidelines for Effective Data Management. Bulletin of the Ecological Society of America, 90(2) 205-214. [HTTP://WWW.NCEAS.UCSB.EDU/NEWS/2009/BORER](http://www.nceas.ucsb.edu/news/2009/borer)

Long Term Ecological Research Network. [HTTP://INTRANET2.LTERNET.EDU/NODE/3248](http://intranet2.lternet.edu/node/3248)

UK Data Archive: Create and Manage Data. [HTTP://WWW.DATA-ARCHIVE.AC.UK/CREATE-MANAGE](http://www.data-archive.ac.uk/create-manage)

Funding agency guidelines

National Endowment for the Humanities. (2011). Available for download [PDF]:

[HTTP://WWW.NEH.GOV/ODH/ODHUPDATE/TABID/108/ENTRYID/162/ANNOUNCING-A-NEW-GRANT-PROGRAM-DIGITAL-HUMANITIES-IMPLEMENTATION-GRANTS.ASPX](http://www.neh.gov/ODH/ODHUPDATE/TABID/108/ENTRYID/162/ANNOUNCING-A-NEW-GRANT-PROGRAM-DIGITAL-HUMANITIES-IMPLEMENTATION-GRANTS.ASPX)

NSF Directorate-level guidance. [HTTP://WWW.LIB.UMICH.EDU/RESEARCH-DATA-MANAGEMENT-AND-PUBLISHING-SUPPORT/NSF-DATA-MANAGEMENT-PLANS](http://www.lib.umich.edu/research-data-management-and-publishing-support/nsf-data-management-plans)

Overview of US federal funding agency policies.

[HTTP://WWW.CDLIB.ORG/SERVICES/UC3/DATAMANAGEMENT/FUNDING.HTML](http://www.cdlib.org/services/uc3/datamanagement/funding.html)

[HTTP://WWW.LIB.UMN.EDU/DATAMANAGEMENT/FUNDING](http://www.lib.umn.edu/datamanagement/funding)

Overview of UK funding agency requirements.

[HTTP://WWW.DCC.AC.UK/RESOURCES/DATA-MANAGEMENT-PLANS/FUNDERS-REQUIREMENTS](http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements)

NIH- Data Sharing Policy.

[HTTP://GRANTS.NIH.GOV/GRANTS/POLICY/DATA_SHARING/DATA_SHARING_GUIDANCE.HTM#EX](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#EX)

Assignment 4: Curation Outreach Presentation (25%)

[DUE: Monday, May 7th by 12pm]

Scenario: You have been invited to speak with a group of researchers in your service community about the role and importance of data curation in the research process. The topic of data curation will be relatively new for most of your audience.

Develop a presentation that covers what you think are the most important things for researchers to understand about data curation, as data producers and potential data providers.

Deliverables:

Prepare a set of slides for a 15-minute talk (no more than 10 slides) and a 1-page handout to be given to attendees.

Your audience may be broadly or narrowly conceived, but the title slide should clearly indicate what audience you are addressing (for example, College of Education, humanities faculty and graduate students, chemists in private industry, ecological field researchers, qualitative social scientists, Economics Department, etc.). Your presentation should draw on the topics and materials from the course as a whole, but you may also wish to emphasize a particular area of relevance to the group, (data formats, documentation and metadata, data sharing, archives for preservation, re-use), dispel common misconceptions of data curation, introduce a valuable tool, or promote a set of services.

The handout should include key points from the presentation and a selected set of 6-10 resources for consultation. Half of these resources can come from the course content, but half must be new materials identified for your audience.

While these presentations will not be given live during class, the materials you prepare should be professional and suitable for sharing with potential employers as part of a portfolio.

Your submission should include a narrated presentation and the 1-page handout. Please use APA citation style.

Please use the following file naming convention: [last name]_A[assignment #]_DCL12.docx (e.g. Sacchi_A4_DCL12.docx)

Additional details on how to create a narrated slide set will be posted to the Forum.

Grading rubric:

- Content of the presentation and handout – high value, succinct, and compelling content for the audience
- Selection of resources – authoritative, accessible, appropriate for the audience
- Design of slides and handout- clean visual presentation, clear and precise language, helpful use of color, images, icons, etc.

Class Schedule

1/19 Week 1- Introduction

1/26 Week 2- Data Curation Landscape

1. Smith, M. (2010). *Communicating Data: New Roles for Researchers, Publishers and Libraries*, CrossRef Annual Meeting. Available: <http://www.youtube.com/watch?v=PoswGVbM8QA>
2. Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4). Available: <http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html>
3. *Uhlir, P. F. (2010). Information gulags, intellectual straightjackets, and memory holes: Three principles to guide the preservation of scientific data. *Data Science Journal*, 10, 1-5.

Background for Assignment 1:

National Science Board. (2005). The elements of the digital data collections universe. In *Long-lived Digital Data Collections*. Pages 17-23, and Appendix D, pages 73-78.

Available: <http://www.nsf.gov/pubs/2005/nsb0540/>

Optional Reading:

1. Westra, B., Ramirez, M., Parham, S. W., & Scaramozzino, J. M. (2010). Selected Internet Resources on Digital Research Data Curation. *Issues in Science and Technology Librarianship*, (63), 9. Available: <http://www.istl.org/10-fall/internet2.html>

2/2 Week 3- Contexts, Concepts, and Models

1. *Parsons, M. A., & Fox, P. A. (submitted 2011). Is data publication the right metaphor? *Data Science Journal*. Available: http://dl.dropbox.com/u/546900/parsons_fox_metaphor_dsj_open.docx
- Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, 57(2), 280-299.
2. Flanders, J. & Muñoz, T. (2011). An introduction to humanities data curation. *DH Curation Guide: a community resource guide to data curation in the digital humanities*. Available: <http://guide.dhcurator.org/intro/>
3. Gold, A. (2010). Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. *Office of the Dean (Library)*. Available: http://digitalcommons.calpoly.edu/lib_dean/27/

Optional readings:

1. Ball, A. (2010). Data lifecycles. In *Review of the State of the Art of the Digital Curation of Research Data*. Project Report. Bath: University of Bath, (ERIM Project Document erimirep09u03ab12). Available: <http://opus.bath.ac.uk/19022/>

2. Association of Research Libraries. (2010, August). E-Science and Data Support Services: A Study of ARL Member Institutions. Available: www.arl.org/bm~doc/escience_report2010.pdf

2/9 Week 4- Data Practices of Researchers (guest speaker: Suzie Allard, University of Tennessee)

1. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. *PloS ONE*, 6(6), e21101. doi:10.1371/journal.pone.0021101
Available: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101>
2. Cragin, M.H., Palmer, C.L., Carlson, J.R., & Witt, M. (2010). Data Sharing, Small Science, and Institutional Repositories. *Philosophical Transactions of the Royal Society A*, 368(1926), 4023-4038.
3. *Babeu, A. Classics, "Digital Classics" and issues for data curation. *DH Curation Guide: a community resource guide to data curation in the digital humanities* Available: <http://guide.dhcuration.org/research-practices/classics/>.
4. Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW)*, 19(3-4), 355-375. doi:10.1007/s10606-010-9117-8

Optional readings:

1. Research Information Network. (2008). *To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs*. Main Report and Annex: Detailed Findings for the Eight Research Areas. See, in particular, the "Classics" case study. Available: <http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs>

See also RIN case studies:

"Reinventing research? Information practices in the humanities." Available:

<http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities>

2. Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing Data Curation Profiles. *The International Journal of Digital Curation*, 4(3), 93-103. doi:10.2218/ijdc.v4i3.117. Available: <http://ijdc.net/index.php/ijdc/article/view/137>
3. Witt, M., & Carlson, J. (2007). Conducting a Data Interview. *Libraries Research Publications*. Available: http://docs.lib.purdue.edu/lib_research/81
4. Data Curation Profiles Toolkit. Available: <http://www4.lib.purdue.edu/dcp/>

2/16 Week 5- Building Data Collections

1. Berman, F. (2010). We need a research data census. *Commun. ACM*, 53(12), 39–41. doi:10.1145/1859204.1859220. Available: <http://cacm.acm.org/magazines/2010/12/102121-we-need-a-research-data-census/fulltext>
2. Faundeen, J. L. and Oleson, L. R. (2007). "Scientific Data Appraisals: The Value Driver for Preservation Efforts" *PV 2007 International Conference*, 9–11 October 2007, DLR, Oberpfaffenhofen/Munich, Germany. Available: www.pv2007.dlr.de
3. Gutmann, M., Schürer, K., Donakowski, D., & Beedham, H. (2004). The selection, appraisal, and retention of social science data. *Data Science Journal*, 3(0), 209–221. Available: http://www.jstage.jst.go.jp/article/dsj/3/0/209/_pdf
4. *Newton, M., Miller, C., & Bracke, M. (2011). Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Management*. 36(1), 53–67.
5. Palmer, C.L., Weber, N.M., & Cragin, M.H. (2011). The analytic potential of scientific data: Understanding re-use potential. In *Proceedings of the American Society for Information Science and Technology Annual Meeting*, October 9–13, New Orleans, LA.

Optional readings:

1. Whyte, A. & Wilson, A. (2010). How to Appraise & Select Research Data for Curation. *Digital Curation Centre, Australian National Data Service*. Available: <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>
2. Inter-university Consortium for Political and Social Research (ICPSR). 'Selection and appraisal'. Available: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/selection.jsp>

Also read:

"Collection Development Policies", available:

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/org/policies/colldev.jsp>

"Appraisal Criteria". Available:

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/appraisal.jsp>

2/23 Week 6- Representation Levels

1. Sperberg-McQueen, M., & Burnard, L. (2002). A Gentle Introduction to XML. *TEI P4: Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/release/doc/tei-p4-doc/html/SG.html>. [Browse]

2. Lynch, C. (1999). Canonicalization: A fundamental tool to facilitate preservation and management of digital information. *D-Lib Magazine*, 5(9).
<http://www.dlib.org/dlib/september99/oglynch.html>
3. *Renear, A., & Dubin, D. (2003). Towards identity conditions for digital documents. *Proceedings of the 2003 international conference on Dublin Core and metadata applications: supporting communities of discourse and practice—metadata research & applications* (p. 19)
<http://dcpapers.dublincore.org/ojs/pubs/article/view/746/742>
4. Tillett, B. (2004). What is FRBR? A conceptual model for the bibliographic universe. *Library of Congress Cataloging Distribution Service* (Vol. 25). <http://www.loc.gov/cds/downloads/FRBR.PDF>
5. Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *Intelligent Systems and Their Applications, IEEE*, 14(1), 20–26.
<http://www.cs.umbc.edu/771/papers/chandrasekaranetal99.pdf>
6. Lowe, B. (2009). DataStaR: Bridging XML and OWL in science metadata management. *Metadata and Semantic Research*, 141–150.
<http://www.springerlink.com/content/qo825vj78ul38712>

Optional readings:

1. Yu, A. C. (2006). Methods in biomedical ontology. *Journal of Biomedical Informatics*, 39(3), 252–266. <http://www.sciencedirect.com/science/article/pii/S1532046405001310>
2. Madin, J. S., Bowers, S., Schildhauer, M. P., & Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends in Ecology & Evolution*, 23(3), 159–168.
<http://www.sciencedirect.com/science/article/pii/S0169534708000384>
3. Doerr, M., & Stead, S. (2009). The CIDOC CRM, a Standard for the Integration of Cultural Information. CRM tutorial presentend at the Imperial College, UK, May 22, 2009. http://www.cidoc-crm.org/docs/crm_for_imperial_2009.ppt.

3/3 Week 7- Interpretation Levels (LEEP weekend)

1. Parkinson, C., Ward, A., & King, M. (2006). Earth science reference handbook. NASA, Washington, DC. http://eosps0.gsfc.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf. [Read section Level Definitions for EOS SDPs, pp 31–32]
2. Renear, A. H., Dolan, M., Trainor, K., & Cragin, M. H. (2009). Towards a cross-disciplinary notion of data level in data curation. *Proceedings of the American Society for Information Science and Technology*, 46(1), 1–8. <http://onlinelibrary.wiley.com/doi/10.1002/meet.2009.14504603103/full>

3. MLA (2007/2011). Guidelines for Editors of Scholarly Editions, Committee on Scholarly Editions, Modern Language Association. http://www.mla.org/cse_guidelines. [Browse]
4. Freire, J., Koop, D., Santos, E., & Silva, C. T. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 11–21. <http://vgc.poly.edu/~juliana/pub/freire-cise2008.pdf>
5. Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *ACM Sigmod Record*, 34(3), 31–36. <http://dl.acm.org/citation.cfm?id=1084812> [Read introductory sections, browse remainder.]
6. *Küster, M., Ludwig, C., Al-Hajj, Y., & Selig, T. (2011). TextGrid provenance tools for digital humanities ecosystems. *Digital Ecosystems and Technologies Conference (DEST), 2011 Proceedings of the 5th IEEE International Conference on* (pp. 317–323). [available online through the LIS Virtual Library]

Optional readings

1. Gil, Y., Cheney, J., Groth, P., Hartig, O., Miles, S., Moreau, L., & Da Silva, P. P. (2010). Provenance XG final report. *Final Incubator Group Report*, (December, 2010). <http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/> [Browse]

3/8 Week 8- Models and Systems

1. CCSDS (2002). Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1, Blue Book. <http://public.ccsds.org/publications/archive/650xob1.pdf>. [Sections 1.1 and 2; section 4.]
2. Farquhar, A., & Hockx-Yu, H. (2008). Planets: Integrated service for digital preservation. *Serials: The Journal for the Serials Community*, 21(2). <http://www.ijdc.net/index.php/ijdc/article/download/45/31>
3. Sharpe, R. (2009). PLANETS Conceptual Data Model. Planets report. (Unpublished draft). [in private circulation]
4. *Duerr, R. E., Downs, R. R., Tilmes, C., Barkstrom, B., Lenhardt, W. C., Glassy, J., Bermudez, L. E., et al. (2011). On the utility of identification schemes for digital earth science data: an assessment and recommendations. *Earth Science Informatics*, 1–22. <http://dx.doi.org/10.1007/s12145-011-0083-6>

3/15 Week 9 (guest speaker: Ruth Duerr on Earth Science)

3/22 (Spring Break)

3/29 Week 10- Costs & Benefits; Data Management Planning

(guest speaker: Sarah Shreeves, University of Illinois Libraries)

1. Beagrie, Charles. (2011, July). Users Guide for Keeping Research Data Safe: Assessing Costs' Benefits of Research Data Management, Preservation, and Re-use (Version 2). *Charles Beagrie Ltd and the Higher Education Funding Council for England*. Available: <http://beagrie.com/krds.php>
2. American Research Libraries (ARL) Guide for Research Libraries: The NSF Data Sharing Policy. (2010). Available: <http://www.arl.org/rtl/eresearch/escien/nsf/index.shtml>
3. *Brown, C. T. (2010, May 17). My data management plan -- a satire. Available: <http://ivory.idyll.org/blog/may-10/data-management.html>
4. Sallans, A. and Donnelly, M. (2011). DMP Online and DMPTool: Different strategies towards a shared goal. Paper presented at the 7th International Digital Curation Conference, December 5-7, Bristol, UK.
 [DMP Online] Digital Curation Centre (DCC) Data Management Planning online.
 Available: <https://dmponline.dcc.ac.uk/>

 [DMPTool] California Digital Libraries (CDL) Data Management Plans Online. (2011).
 Available: <http://www.cdlib.org/uc3/datamanagement/dmpo.html>. See also "DMP Tool".
 Available: <https://dmp.cdlib.org/>

Optional readings:

1. Association of Research Libraries (ARL) Data Management Planning. (2011). Available: <http://www.arl.org/rtl/eresearch/escien/nsf/nsfresources.shtml>
 [Details on Discipline-based repositories] University of Minnesota Libraries. Available: <http://www.lib.umn.edu/datamanagement/DMP/example>

 [Details on Federal funding agency requirements] Michigan State University Libraries (MSU). (2011). Lifecycle Data Management Planning.
<http://www.lib.msu.edu/about/diginfo/ldmp.jsp>

4/5 **Week 11- Policy and regulatory environment**

1. Jones, S. (2009). A report on the range of policies required for and related to digital curation, version 1.2, (DCC, Glasgow, March 2009). Available: <http://hdl.handle.net/1842/3371>
2. Green, A., Macdonald, S., & Rice, R. (2009). Policy-making for Research Data in Repositories: A Guide (v.1.2.). *Data Information Specialists Committee-UK*. Available: www.disc-uk.org/docs/guide.pdf
3. *Stodden, V. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science & Engineering*, 11(1), 35-40. doi:10.1109/MCSE.2009.19
4. Lehmberg, T., Rehm, G., Witt, A., & Zimmermann, F. (2008). Digital text collections, linguistic research data, and mashups: notes on the legal situation. *Library Trends*, 57(1), 52-71.

Optional readings:

1. Wilbanks, J. (2010, June 2). Open data and Creative Commons: It's about scale... Available: http://scienceblogs.com/commonknowledge/2010/06/open_data_and_creative_commons.php
2. Science Commons. Protocol for Implementing Open Access Data. Available: <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>
3. Open Knowledge Foundation. Open Definition. Available: <http://opendefinition.org/okd/>
4. Levine, M. "Policy, practice, and law". *DH Curation Guide: a community resource guide to data curation in the digital humanities*. Available: <http://guide.dhcuration.org/legal/policy/>

4/12 **Week 12- Infrastructure** (guest speaker: Mark Evans, Tessella)

1. Wilson, J. A. J., Martinez-Urbe, L., Fraser, M. A., & Jeffreys, P. (2011). An Institutional Approach to Developing Research Data Management Infrastructure. *International Journal of Digital Curation*, 6(2). doi:10.2218/ijdc.v6i2.203. Available: <http://www.ijdc.net/index.php/ijdc/article/view/198>
2. *Goddard, A., Wilson, N., Cryer, P., & Yamashita, G. (2011). Data hosting infrastructure for primary biodiversity data. *BMC Bioinformatics*, 12 (Suppl 15), S5. Available: <http://www.biomedcentral.com/1471-2105/12/S15/S5>
3. UC Curation Center (UC3). Curation Micro-Services. Available: <http://www.cdlib.org/services/uc3/curation/>

See also link at bottom ('Background Information') to: UC Curation Center/California Digital Library. (2010). Merritt: An Emergent Micro-Services Approach to Digital Curation Infrastructure. *[opens a PDF]*

4. Data Conservancy Infrastructure case study. [working draft]

Optional readings:

1. "Getting started with Fedora." (version 3.5) Available:
<https://wiki.duraspace.org/display/FEDORA35/Getting+Started+with+Fedora>
2. "iRODS Overview." Available: https://www.irods.org/index.php/iRODS_Overview
3. DRAMBORA Interactive. "DRAMBORA: About." Available: <http://www.repositoryaudit.eu/about/>
See following documents for sample of Toolkit [see Moodle attachments]:
 DRAMBORA Toolkit (full text and explanation of Toolkit)
 DRAMBORA MS Word templates
 DRAMBORA Excel templates
4. Center for Research Libraries (CRL) and Online Computer Library Center (OCLC). (2007).
 "Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)."
 http://www.crl.edu/sites/default/files/attachments/pages/trac_o.pdf
 See pages 51-72 for sample of Checklist criteria

4/19 Week 13- guest speaker (Julia Flanders on Humanities; Brown University Women Writers Project)

4/26 Week 14 (Final class)

LIS586A Digital Preservation
Spring 2012
Thursday, 9:00 - 11:50 a.m.

COURSE DESCRIPTION

This course examines the current problems with and approaches to digital preservation that are fundamental to the long-term accessibility of digital materials. We will examine the range of current research problems, along with emerging methods and tools, and assess a variety of organizational scenarios to plan and implement a preservation plan. Topics will include basic information theory, preservation of complex digital objects; standards and specifications; sustainability and risk assessment; authenticity, integrity, quality control, and certification; and management of preservation activities.

NB: All course readings are to be completed before lecture in the week in which they are listed.

COURSE INSTRUCTOR

Jerome McDonough
Assistant Professor, GSLIS
jmcdonou@uiuc.edu
(217) 244-5916
202 LIS Building, MC-493
501 E. Daniel Street
Champaign, IL 61820
Office Hours: Wednesday 10:00 - 12:00 or by appointment

ASSIGNMENTS

Class Participation: 20%

Participation will be evaluated based on your contributions both during class and on the webboard. Please note that the quality of your contributions counts. Informed comments, thoughtful exchanges with your fellow students and demonstrating your interest and engagement with the assignments, reading materials, and general topics are the sure path to a good participation grade.

Assignment 1: 25%

Analytic Essay

Due: Variable

The analytic essay assignment is intended to allow you to further explore one of the course topics that is of particular interest to you. For this assignment, you will select and read at least 3 additional peer-reviewed journal articles on a particular week's topic, and write an essay (8-10 pages, double space) which will critically analyse and discuss those additional readings; you may also include the assigned readings in your discussion. You decide which week you would like to cover, but please post to the Analytic Essay discussion forum indicating the week for which you'll be writing your essay. The essay is due prior to the meeting of that week's class, and should be submitted through Moodle in MS-Word format and also posted to that week's discussion forum. Citations should follow APA style.

NB: The best essays tend to be those which are making an argument, not just summarizing the readings. Try to develop a position of your own based on the readings and advocate for it.

Assignment 2: 30% (3 sections, 10% each)**Preservation Format Identification & Description****Due: Part 1 - Feb. 16 (In Class), Part 2 - Feb. 23 (In Class), Part 3 - March 8**

This assignment will focus on identifying file formats, determining technical properties of files in particular file formats, and recording that information using current preservation metadata standards such as PREMIS. The assignment will have 3 parts, the first two of which will be completed in class, and the second of which will be completed over two weeks (with time allowed in-class for the work).

Part 1 (Feb. 16)

Identification of file formats is key to knowing the contents of your digital repository and determining an appropriate preservation strategy. In this assignment, you will be provided with a set of 8 files lacking any obvious format identification (such as a filename extension). You will be shown how to download, install and use the DROID file format identification software from the PRONOM project in the UK as well as how to use the Unix "file" utility to identify file formats. Using these tools, you will generate a list of file formats and checksum values for the provided files and submit them at the end of class.

Part 2 (Feb. 23)

In addition to identifying file formats, you need to determine technical information about the files (such as sampling bit depth, color space, dimensions, etc.). In this exercise, you will be shown how to use a variety of tools to obtain technical information about text, image, audio and video files. Using those tools, you will need to obtain a set of technical information on each of the files in the preceding assignment.

For text files, you will need to supply the following pieces of information:

File Size

Language (as ISO 639-2 code)

Character size (number of bytes per character)

Character set

Markup Basis (is the file XML, SGML, GML, etc.)

Markup Language (what specific Markup Language does the file use, e.g., HTML, TEI, DocBook)

For Image Files:

File Size

Byte Order

Compression

Image Width

Image Height

Color Space

For Audio Files

File Size

Duration

Number of Channels (e.g., mono, stereo, 5.1 surround sound)

Sampling Rate

Bit Depth

Compression

For Video Files

File Size

Duration
Number of tracks and types (e.g., 1 video, 2 audio, 1 caption)
Frame size
Frame Rate
Compression

Be aware that not every file will exhibit each of the above characteristics. Certain image files, for example, will not use compression, or will not use multiple bytes per sample and so byte order is not relevant.

Part 3 (March 8)

Of this assignment, you will use the information you obtained in the previous two parts of the assignment to create a metadata record for one of the files in the set using the PREMIS metadata standard. You will submit this PREMIS record through the Moodle system.

Assignment 3: 25% [group]

Choose either A. Preservation Planning or B. Planning Grant Application

Due: May 3

For your final assignment, you will form groups of 3 to 5 people per team, and work on either option A (Preservation Planning) or B (Planning Grant Application). Each group will need to submit by March 15th a preliminary work plan which includes: 1. the names of the members of your group, 2. identification of one individual who'll be serving as team leader for your group, 3. an indication of whether your group is working on option A or option B, and if option B, the digital collection your group wishes to work with. Further details on each assignment option are below.

A. Preservation Planning Assignment

Your group consists of professionals working in the Rare Books and Manuscripts Library of the University of Illinois which has just been given the archives of a very famous and important writer who is an alumnus. The addition of this archive will greatly enhance the school and the library's stature in the field of contemporary studies. In addition to manuscripts and other papers there is a large collection of electronic files that have been received on several kinds of media including hard drives, floppies, CDs and DVDs. The University has an IT department that runs a campus-wide network which is backed-up nightly but has no pretensions of being a repository. There is also IDEALS which "collects, disseminates, and provides persistent and reliable access to the research and scholarship of faculty" but which is not designed to maintain an archive of this nature. The University administration realizes that it must take some action to make this collection accessible and to preserve it and, further, that it will need to plan on more and more electronic resources of this kind in the future. Because they have all taken the course in Digital Preservation, the people in your group have been chosen to form a committee to begin the planning process. Your charge is: With a mind to the continued sustainability determine the immediate disposition of the collection (including treatment of original media/computers); recommend how it should be integrated into the existing collections; make recommendations on how they can be accessed now and in the future; and determine policy changes are needed to accommodate the administration and preservation of electronic resources.

You will be given a set files for this exercise that will be examples of all the formats included in the collection. You will deliver a white paper that will explain your recommendations to the Library and University administration with suitable explanations supported by references. Your recommendations should be prioritized so that the administration can adjust their implementation to existing resources but they should also make clear what actions should be taken, what standards should be met, what resources should be assigned, etc.

B. Planning Grant Application

Your team will collaborate on writing a grant proposal to the National Endowment for the Humanities' Preservation and Access: Humanities Collections and Resources program, seeking funds to evaluate the preservation needs of a digital collection (<http://www.neh.gov/grants/guidelines/HCCR.html>). The objective of the grant will be to secure funds to assist in “preserving and improving access to humanities resources in 'born digital' form.” You will collaborate with a partner organization to gather information on the current environment, analyze the preservation needs and the existing preservation situation, present findings, and propose recommendations for a grant-funded project. Your final submission should follow the outlines of the call for proposal which includes a one page abstract, a table of contents, a 10-15 page (single-spaced) project narrative, an itemized budget for project expenses, and supporting documentation.

Potential Partner Organizations:

Stanford University Libraries / The Stephen Cabrinety Collection in the History of Microcomputing. See <http://library.stanford.edu/depts/hasrg/histsci/index.htm>

University of Maryland / Deena Larsen Collection. See <http://www.mith.umd.edu/larsen/>

OR

A Collection of your choosing (Clear your choice with Prof. McDonough before proceeding).

CALENDAR

January 15 - January 21

Introduction

Task Force on Archiving of Digital Information (May 1996). Preserving Digital Information, Report of the Task Force on Archiving of Digital Information. Donald Waters and John Garrett, Eds. The Commission on Preservation & Access and the Research Libraries Group. Available at: <http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf>

Task Force on the Artifact in Library Collection (November 2001). The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections. Stephen G. Nichols & Abby Smith, Eds. Council on Library & Information Resources. Available at: <http://www.clir.org/pubs/reports/pub103/pub103.pdf>

January 22 - January 28

Information Theory

Pierce, John R. (1980). An Introduction to Information Theory: Symbols, Signals and Noise. 2nd Revised Edition. New York: Dover Publications, Inc.

NB: Chapters II - VI only -- available through UIUC e-Reserves

Suggested Optional Readings:

Shannon, Claude (1948). "A mathematical theory of communication." Bell System Technical Journal 27. pp. 379–423, 623–656. Available at: <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>

January 29 - February 4

Archival Theory & Diplomats

Gilliland-Swetland, Anne J. (Feb. 2000). Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment. Council on Library & Information Resources. <http://www.clir.org/pubs/reports/pub89/pub89.pdf>

The InterPARES Project (2001). The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project. <http://www.interpares.org/book/index.cfm>

Duranti, Luciana (2001). "The Impact of Digital Technology on Archival Science" in *Archival Science* 1(1), pp. 39-55. Available through UIUC electronic resources.

Duranti, Luciana (Fall 2009). "From Digital Diplomatics to Digital Records Forensics." *Archivaria* 68, pp 39-66. Available through UIUC electronic resources.

Suggested Optional Readings:

Mortensen, Preben (Spring, 1999). "The Place of Theory in Archival Practice." *Archivaria* 47, pp. 1-26.

Dollar, C. M. (1992). *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods*. Macerata: University of Macerata Press.

Jenkinson, Hillary (1984). "Reflections of an Archivist" in *A Modern Archives Reader*, Washington, DC: National Archives & Records Service.

February 5 - February 11

Open Archival Information System Reference Model

Consultative Committee for Space Data Systems (August 2009). Reference Model for an Open Archival Information System. CCSDS 650.0-P-1.1. Pink Book. Issue 1.1.

<http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf>

Tansley, Robert, Bass, Mick & Smith, MacKenzie (2004). "DSpace as an Open Archival Information System: Current Status and Future Directions" in *Lecture Notes in Computer Science* 2769, pp. 446-460. Available through UIUC E-Resources.

February 12 - February 18

Preservation Strategies

Thibodeau, Kenneth (July 2002). "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years." In *The State of Digital Preservation: An International Perspective Conference Proceedings, Washington, DC, April 24-25, 2002*. Council on Library & Information Resources, Washington, DC., pp. 4-31. Available at <http://www.clir.org/pubs/abstract/pub107abst.html>

Margaret Hedstrom and Clifford Lampe (Dec. 15, 2001). "Emulation vs. Migration: Do Users Care?" in *RLG DigiNews* 5(6). Available at

<http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2448.html#feature1>

Ippolito, Jon (). "Accommodating the Unpredictable: The Variable Media Questionnaire". In Alain Depocas, Jon Ippolito and Caitlin Jones (Eds.) *The Variable Media Approach: Permanence Through Change*, pp. 47-55. Available at: <http://www.variablemedia.net/pdf/Permanence.pdf>

Computer History Museum. The PDP-1 Restoration Project. Available at: <http://pdp-1.computerhistory.org/pdp-1/index.php>. NB: make sure to listen to the interview with Joe Fredrick and to read the Introduction and all of the subsections under the Restoration heading (Team, Hardware, Software, Input/Output).

Recommended Optional Readings

Rothenberg, Jeff (Jan. 1999). *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington, DC: Council on Library & Information Resources. Available at:

<http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>

Bearman, David (1999). "Reality and Chimeras in the Preservation of Electronic Records" in D-Lib Magazine 5(4). Available at: <http://www.dlib.org/dlib/april99/bearman/04bearman.html>

Granger, Stewart (2000). "Emulation as a Digital Preservation Strategy" in D-Lib Magazine 6(10). Available at: <http://www.dlib.org/ar/dlib/october00/granger/10granger.html>

Holdsworth, David and Paul Wheatly (Aug. 15, 2001). "Emulation, Preservation, and Abstraction". RLG DigiNews 5(4). Available at: <http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2>

Verdegem, Remco & Slats, Jacqueline. "Practical Experiences of the Dutch Digital Preservation Test-Bed." *Vine* 34(2), pp. 56-65.

February 19 - February 25

Data Formats

Lawrence, Gregory W. et al. (2000). Risk Management of Digital Information: A File Format Investigation. Washington DC: Council on Library & Information Resources. Available at: <http://www.clir.org/pubs/reports/pub93/pub93.pdf>

Pearson, David & Webb, Colin (July 2008). "Defining File Format Obsolescence: A Risky Journey." International Journal of Digital Curation 3(1). Available at: <http://www.ijdc.net/index.php/ijdc/article/viewFile/76/44>

Arms, Caroline and Fleischhauer, Carol (2005). "Digital Formats: Factors for Sustainability, Functionality and Quality." Paper presented at IS&T Archiving 2005 Conference, Washington, DC. Available at: http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf

Library of Congress (2006). "Content Categories," a subsection of Sustainability of Digital Formats: Planning for Library of Congress Collections. Available at: http://www.digitalpreservation.gov/formats/content/content_categories.shtml

NB: Look at information for all five content categories of Still Image, Sound, Textual, Moving Image and Web Archive, including sub-pages on 'Quality and Functionality Factors,' 'Preferences in Summary,' and 'Curator's View.' You do not need to explore 'Format description documents' in depth, although you may want to look at one or two.

Suggested optional readings:

Coleman, James and Don Willis (July 1997). SGML as a Framework for Digital Preservation and Access. Washington, DC: Council on Library & Information Resources.

February 26 - March 3

Metadata

Lavoie, Brian F. (Summer 2004). "Preservation Metadata: Challenge, Collaboration, and Consensus" in Microform & Imaging Review 33(3), p. 130-4. Available through E-Reserves.

PREMIS Working Group. Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group. OCLC, Inc. and Research Libraries Group. NB: Only sections 1 & 2 are mandatory; rest are optional.

Blank, Grant & Rasmussen, Kiarsten Boye (2004). "The Data Documentation Initiative," Social Science Computer Review, 22(3), pp. 307-318, DOI: 10.1177/0894439304263144. Available through UIUC E-Journals

Hunter, Jane and Choudhury, Sharmin (April 2006). "PANIC: an integrated approach to the preservation of composite digital objects using Semantic Web services" in *International Journal on Digital Libraries* 6(2). Available through UIUC E-Journals.

Draper, Hal (December 1961). "MS FND IN A LBRY: or, The Day Civilization Collapsed." *The Magazine of Fantasy and Science Fiction* 21, pp. 76-81. Available through UIUC E-Reserves.

Suggested optional readings:

Lazinger, Susan F. (2001). *Digital preservation and metadata: history, theory, practice*. Englewood, CO: Libraries Unlimited.

Data Documentation Initiative Website at <http://www.icpsr.umich.edu/DDI/>

March 4 - March 10

Preservation Systems

Reich, V. & D. S. H. Rosenthal (June 2001). "LOCKSS: A Permanent Web Publishing and Access System." *D-Lib Magazine* 7(6). Available at: <http://www.dlib.org/dlib/june01/reich/06reich.html>

Caplan, P. (2007). "The Florida Digital Archive and DAITSS: A Working Preservation Repository Based On Format Migration." *International Journal of Digital Libraries* 6(4). Available through UIUC Library Online Resources

York, Jeremy (2009). "This Library Never Forgets: Preservation, Cooperation, and the Making of HathiTrust Digital Library." *Archiving 2009: Final Program & Proceedings*. May 4 - 7, 2009, Arlington, VA. Available at: <http://www.hathitrust.org/documents/This-Library-Never-Forgets.pdf>

Woods, A. (2010). "DuraCloud: Preservation Infrastructure in the Cloud." A presentation at the *CNI Spring Membership Meeting, Baltimore, MD, April 2010*. Video of presentation is available at <http://www.youtube.com/watch?v=dSEfZfEdobI> and a PDF of the presentation slides is available at <http://duraspace.org/documents/DuraCloudCNIApril10.pdf>

Optional Readings

If you want a bit of background on cloud computing, see

Lenk, A., M. Klems, J. Nimi, S. Tai and T. Sandholm (2009). "What's inside the Cloud? An Architectural Map of the Cloud Landscape." *Proceedings of the 2009 International Conference on Software Engineering Workshop on Software Engineering Challenges of Cloud Computing*. Available at: <http://portal.acm.org/citation.cfm?id=1564625>

On possible problems with cloud storage, have a look at

Cachin, C., I. Keidar and A. Shraer (June 2009). "Trusting the Cloud." *ACM SIGACT News* 40(2). Available at: <http://portal.acm.org/citation.cfm?id=1556173>

March 11 - March 17

Preservation System Assessment

OCLC Online Computer Library Center and The Center for Research Libraries (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Available at: <http://www.crl.edu/PDF/trac.pdf>

University of Illinois Libraries. *Trustworthy Repositories Audit & Certification: Criteria Checklist for IDEALS*. Available

at: <https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/Internal/TrustworthyRepositoriesAuditChecklist>

NB: Click through on each major section (A, B & C) and you will get a page showing the results for each of the checklist items (including identification of things to do still).

McHugh, Andrew, Perla Innocenti, Seamus Ross & Raivo Ruusalepp (Sept. 20, 2007). "Risk Management Foundations for Digital Libraries: DRAMBORA" in Donatella Castelli & Yannis Ioannidis (Eds.) *Foundations of Digital Libraries - II*. Available at: http://eslking.com/library2/collections/librarianship/digital_libraries/foundations_of_dl.pdf#page=7

Innocenti, Perla & Giuseppina Vullo (2009). "Assessing the Preservation of Institutional Repositories with DRAMBORA: Case Studies from the University of Glasgow." *Bollettino Associazione Italiana Biblioteche* (2). Available at: <http://www.aib.it/aib/boll/2009/0902139.htm>

March 18 - March 24

SPRING BREAK

March 25 - March 31

Data Curation

Moore, Reagan W. (2003). Preservation of Data. San Diego Supercomputing Center Technical Report 2003-06. Available at: <http://www.sdsc.edu/NARA/Publications/data-preservation.doc>

Beagrie, Neil (2006). "Digital Curation for Science, Digital Libraries, and Individuals," *The International Journal of Digital Curation* 1(1), pp. 3-16. Available at: <http://www.ijdc.net/ijdc/article/view/6/5>

Giaretta, David (2007). "The CASPAR Approach to Digital Preservation," *International Journal of Digital Curation* 2(1). Available at: <http://www.ijdc.net/ijdc/article/view/29/32>

Karasti, Helena, Baker, Karen S. & Halkola, Eija (2006). "Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network," *Computer Supported Cooperative Work* 15(4), pp. 321-358. Available through UIUC E-Journals.

Recommended Optional Reading

Beagrie, Charles, Lavoie, Brian & Woollard, Matthew (2010). *Keeping Research Data Safe 2*, a study funded by JISC with support from OCLC Research and the UK Data Archive. London: Higher Education Funding Council for England. Available at: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>

April 1 - April 7

Authenticity, Integrity & Trust

RLG/OCLC Working Group on Digital Archive Attributes (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Research Libraries Group. Available at: <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

Smith, MacKenzie & Moore, Reagan W. (2006). "Digital Archive Policies and Trusted Digital Repositories." Workshop on Digital Curation & Trusted Repositories: Seeking Success, June 15, 2006 in conjunction with the Joint Conference on Digital Libraries (JCDL 2006, June 11-15, 2006 - Chapel Hill, NC, USA). Available at: <http://dspace.mit.edu/handle/1721.1/38868>

Smith, Abby (Summer 2003). "Authenticity and Affect: When is a Watch Not a Watch?" in *Library Trends* 52(1), pp. 172-82. Available through UIUC E-Journals.

Depocas, Alain, Ippolito, Jon and Caitlin Jones (Eds.) (2003). *Variable media approach : permanence through change*. The Solomon R. Guggenheim Foundation and the Daniel Langlois Foundation for Art, Science, and Technology. Available at: <http://www.variablemedia.net/pdf/Permanence.pdf>

Council on Library & Information Resources (May 2000). *Authenticity in a Digital Environment*. Washington, DC: Author. Available at: <http://www.clir.org/pubs/reports/pub92/pub92.pdf>

Additional Optional Readings

Zhao, Jun, Graham Klyne, Guillermo Alvaro Rey, Jose Manuel Gomez-Perez (May 31, 2011). *Workflow Integrity and Authenticity Maintenance Initial Requirements*. Wf4Ever: Advanced Workflow Preservation Technologies for Enhanced Science. Available at: <http://repo.wf4ever-project.org/Content/18/D4.1.pdf>

April 8 - April 14

Evaluation & Value

Rauch, C. and A. Rauber (2004). "Preserving digital media: Towards a preservation solution evaluation metric" in *Digital Libraries: International Collaboration and Cross-Fertilization Proceedings, Lecture Notes in Computer Science* 3334, pp. 203-212. Available through UIUC E-Journals.

Rauch, Carl et al. (2005). "Evaluating preservation strategies for audio and video files" in *Proceedings of the DELOS Workshop on Digital Repositories: Interoperability and Common Services*, May 11-13 2005, Heraklion, Greece. Available at: http://www.dpc.delos.info/private/output/rau_digrep05.pdf

Smith, Abby (May 2003). "Issues in sustainability: Creating value for online users" in *First Monday* 8(5). Available at: http://www.firstmonday.org/issues/issue8_5/smith/index.html

Duranti, Luciana (1994). "The concept of appraisal and archival theory" in *The American Archivist* 57, pp. 328-344. Available through UIUC e-Reserves.

Currall, James and McKinney, Peter (April 2006). "Investing in Value: A Perspective on Digital Preservation" in *D-Lib Magazine* 12(4). Available at: <http://www.dlib.org/dlib/april06/mckinney/04mckinney.html>

Optional Readings

Rosenthal, David (2011). *Modeling the Economics of Long-Term Storage*. Blog post. Available at: <http://blog.dshr.org/2011/09/modeling-economics-of-long-term-storage.html>

April 15 - April 21

Digital Preservation & The Law

Charlesworth, Andrew (2003). *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: A study undertaken for the JISC and Wellcome Trust*. The Joint Information Systems Committee & The Wellcome Trust. Available at: http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf

United States Code, Title 17, Chapter 1. Available at: <http://www.copyright.gov/title17/92chap1.html>

Council on Library & Information Resources and The Library of Congress (2010). *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age*. Commissioned for and sponsored by the National Recording Preservation Board of the Library of Congress. Washington, DC:

Council on Library & Information Resources. Available at:
<http://www.clir.org/pubs/reports/pub148/pub148.pdf>. READ CHAPTER 4 ONLY.

Library of Congress, Copyright Office (2006). "Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control," Federal Register 71(227), pp. 68472-68480. Available at:
<http://www.copyright.gov/fedreg/2006/71fr68472.html>

Creative Commons (2010). Licenses - Creative Commons. Available at:
<http://creativecommons.org/about/licenses/>

Hirtle, Peter (2010). Copyright Term and the Public Domain in the United States, 1 January 2010. Available at: <http://copyright.cornell.edu/resources/publicdomain.cfm>

April 22 - April 28

Reports from the Field

Rusbridge, Chris (Feb. 2006). "Excuse me... Some digital preservation fallacies?" *Ariadne* 46. Available at: <http://www.ariadne.ac.uk/issue46/rusbridge/?ref=cia-team.com>

Caplan, Priscilla (July 2007). "The Florida Digital Archive and DAITSS: A Working Preservation Repository Based on Format Migration." *International Journal on Digital Libraries* 6(4). Available through UIUC E-Resources.

Bagdanov, Andy et al. (May 2009). Fedora Commons 3.0 vs. DSpace 1.5: Selecting an Enterprise-Grade Repository System for FAO Of the United Nations. Available at:
<http://smartech.gatech.edu/handle/1853/28417>

Cathro, Warwick, Webb, Colin & Whiting, Julie (2001). Archiving the Web: The PANDORA Archive at the National Library of Australia. *Preserving the Present for the Future Web Archiving Conference, Copenhagen, 18-19 June, 2001*. Available at:
<http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1314/1600>

Case Study in Data Curation: Crystallography

INTRODUCTION and BACKGROUND

Crystallography is the science of determining the arrangement of atoms in solid materials; this is generally referred to as "solving a structure." The specific technique of X-ray diffraction is carried out very widely for solving single, small crystal structures, particularly in academic settings (as the technology for this is more easily accessible). There are several problems related to sharing and access to crystallographic data. Based on research by Allen (2002, 2004), for example, Coles et. al. (2005) estimate that, "1.5 million structures have been determined in research laboratories worldwide," yet "less than 20% of data generated in crystallographic work is reaching the public domain. ... For the research chemist just 300,000 crystal structures are available in subject specific databases that have harvested their content from the published literature" (p. 3). Providing access to more of these data would support research in several fields.

Crystallography is an experimental research field in which the data are often generated and processed as a service to support the research of others, such as chemists and materials scientists. While deposit of a final form of the data is required for publication of structures, many of the structures that are (and have been) solved are never published. In addition, metadata or other information associated with the crystal structure (e.g. provenance) is not always made available. There is a concern about file formats, as data are produced in a range of formats which are generally based on the instrument technology or the software applications used for processing the data.

Raw crystallographic data are generally considered to be the diffraction images that contain information that must be decoded in order to obtain a reduced data file (ascii text) from which the results may be derived. These images produce large data sets that can be as large as a gigabyte. These data have enduring value, though in many labs they are rarely kept beyond the short term. Processed data types include the amplitude measures of atoms in the data space, as well as "corrected" measurements based on current theory. It is these "corrected" numbers that are generally used to generate the result published in a paper. Finally, CIF data files are written based on a standard maintained by the International Union of Crystallography (<http://www.iucr.org/>); these are generally required for publication in chemistry journals though there is some variability in terms of completeness. These are fairly small text files that hold the data on the location, characteristics and pattern of atoms. Standards that might apply to these data include CML (Chemical Markup Language), an open source XML tool for encoding molecular information.

A significant component to acquiring these data for shared or public access is the particular circumstances around ownership of the data. Since the data are generated and processed in a service model - that is, the structures are most often solved by a crystallographer for a chemist, for example - ownership of the data can be problematic. There is a need for crystallographers to negotiate with chemists (or others) at the time that the data are transferred to them, for how and when the data might be released for public use. If that researcher decides not to publish the resolved crystal structure, the crystallographer may be able to release the data themselves.

CURRENT SITUATION / SCENARIO: Several researchers on campus would like to undertake a study comparing the results from previously generated data with the results from the application of a new technique. There is not currently a way for them to access this data. The local crystallographer has asked for some data management support. In addition, the University would like to see data for doctoral dissertations made available. Let's assume that the university library runs an Institutional Repository.

QUESTIONS for STUDENTS:

First, what other information about this field might be helpful in planning support or services? Do we need to know more from the local crystallographer?

How might we approach this opportunity? (And, where would we begin?)

How might we approach the distribution and management of the various data? (For example, which data might be appropriate for ingest into the IR?)

How might we approach the storage and access to the data related to the dissertations?

ACKNOWLEDGEMENTS

Scott Wilson, Ph.D., Director of the George L. Clark X-ray Facility and the 3M Materials Chemistry Laboratory here at UIUC, and Roxy Wilson, Ph.D., and Lecturer in Chemistry at UIUC both spent a Monday afternoon discussing this case and providing critical background on the data processing activities involved in x-ray crystallography. Dr. Simon J. Coles, the manager of the UK National Crystallography Service at the School of Chemistry at the University of Southampton, provided also expert advice on the descriptions of crystallography data as well as intellectual property issues.

REFERENCES

Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.*, B58, (2002), 380-388.

Allen, F. H. High-throughput crystallography: the challenge of publishing, storing and using the results. *Crystallography Reviews*, 10 (2004), 3-15.

Coles, Simon, Frey, Jeremy, Hursthouse, Michael, Light, Mark, Carr, Leslie, DeRoure, David, Gutteridge, Christopher, Mills, Hugo, Meacham, Ken, Surridge, Mike, Lyon, Liz, Heery, Rachel, Duke, Monica and Day, Michael (2005). The 'end to end' crystallographic experiment in an e-Science environment: From conception to publication. In, UK e-Science All Hands Meeting 2005, Nottingham, UK, 19-23 Sep 2005. Southampton, UK. Available: <http://eprints.soton.ac.uk/17454/>. Accessed Sept. 30, 2007.

Coles, Simon J., Jeremy G. Frey, Michel B. Hursthouse, Mark E. Light, Andrew J. Milsted, Leslie A. Carr, David DeRoure, Christopher J. Gutteridge, Hugo R. Mills, Ken E. Meacham,

Michael Surridge, Elizabeth Lyon, Rachel Heery, Monica Duke, and Michael Day. (2006). An E-Science Environment for Service Crystallography-from Submission to Dissemination. J. Chem. Inf. Model., 46 (3), 1006 -1016, 2006. 10.1021/ci050362w S1549-9596(05)00362-1

CML - http://wwmm.ch.cam.ac.uk/wikis/wwmm/index.php/Chemical_Markup_Language

See also: <http://www.ch.ic.ac.uk/rzepa/chimeral/documents/FAQ/FAQ.html>

International Union of Crystallography: <http://www.iucr.org/>

See also the CIF checker: <http://checkcif.iucr.org/>

Appendix D: Recruitment Materials

The following is an example of the recruitment letters DCEP sent to targeted institutions. Each recruitment packet also included a poster and a flyer (included on the following pages). In later semesters, we created a second flyer (also included) used for a variety of recruitment activities.

Dear [McNair Scholars Program Director],

Our Data Curation Education Program (DCEP) specialization has launched with great success, and now provides a cutting-edge career-track specialization within the Master of Science degree program at the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. Students who enter our program will receive the benefit of paid internship opportunities, the option to apply for fellowships, and the additional advantages of attending a #1 ranked library and information science school. We have identified your department as a potential resource to recruit future students, and would appreciate your help in advertising our program by posting the enclosed flyers around your campus to make students aware of these opportunities.

Digital data production is seeing exponential growth, and scholarly and research-oriented organizations are expending time and resources to support the management of institutionally-generated data. As these efforts develop, organizations will require new kinds of expertise for the proper appraisal, management, and access to data for long-term use. Data curation is the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time. At GSLIS, students in the Data Curation concentration acquire training in general librarianship and information science, and also develop specific skills through curriculum intended for students interested in technical training for work within academia, industry and museums. Students have the opportunity to learn about the theories and practices of library and information science with an emphasis on knowledge representation, digital preservation and archiving, and data standards and policy. Students learn to create and maintain data collections, manage data systems, and evaluate and apply data and metadata standards for varied uses across the sciences and humanities.

Our Data Curation specialization is producing the next generation of experts in data management and care. Once students have acquired technical training through the Data Curation concentration, they will be prepared to apply their skills in

- **Digital Libraries**
- **Data Repositories**
- **Research Laboratories**

For more information on the data curation concentration, please visit:

http://www.lis.uiuc.edu/programs/ms/data_curation.html

Sincerely,

Rae-Anne Montague
Assistant Dean, Student Affairs

Will today's digital information be accessible tomorrow?

MASTER OF SCIENCE CONCENTRATION IN

DATA CURATION


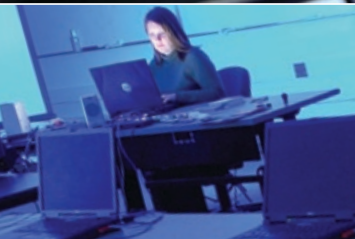
www.lis.uiuc.edu/programs/ms/data_curation.html



Every aspect of our lives has been touched by the digital revolution. We share our pictures on Flickr, watch videos on YouTube, blog our daily lives. We manage our healthcare, finances, and education online. In science, industry, and academia we are generating data and new kinds of information at an exponential rate.

GSLiS

Graduate School of Library and Information Science



Information professionals are needed to build and preserve our digital world. Our new master's degree concentration in data curation will prepare you to develop and use the tools needed to preserve and share digital content within a variety of research communities. You will explore issues such as digital preservation and archiving, data standards, and policy. As a professional in the field, you will ensure that in 5 years or 50 years, today's information remains secure, stable, and easily retrievable.

Graduate School of Library and Information Science

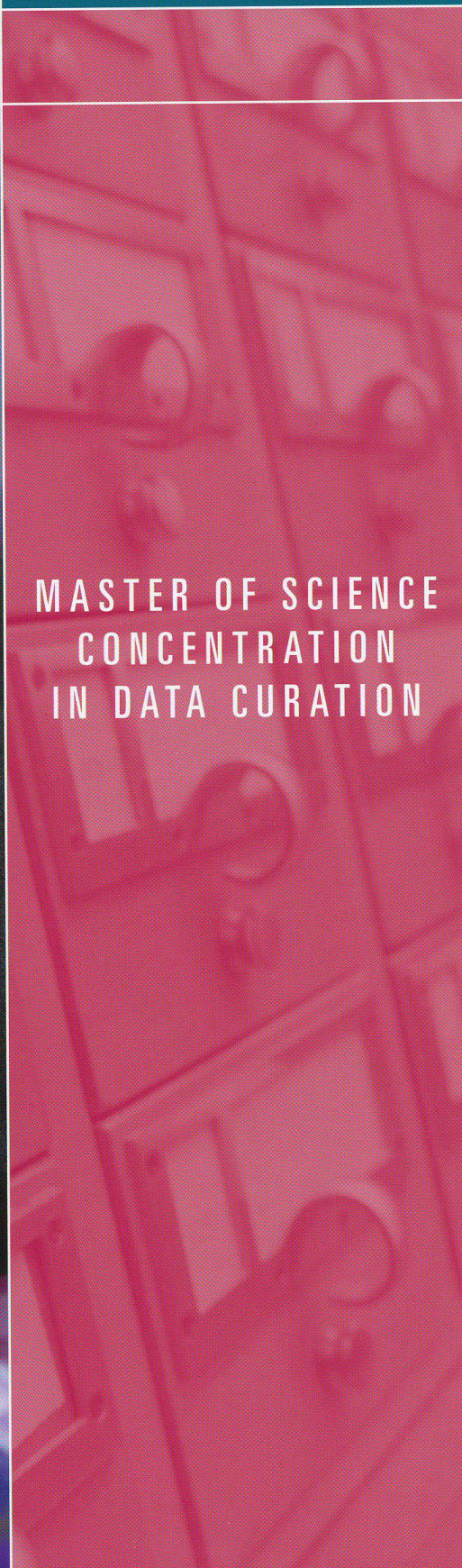
GSLiS



University of Illinois at Urbana-Champaign



MASTER OF SCIENCE
CONCENTRATION
IN DATA CURATION

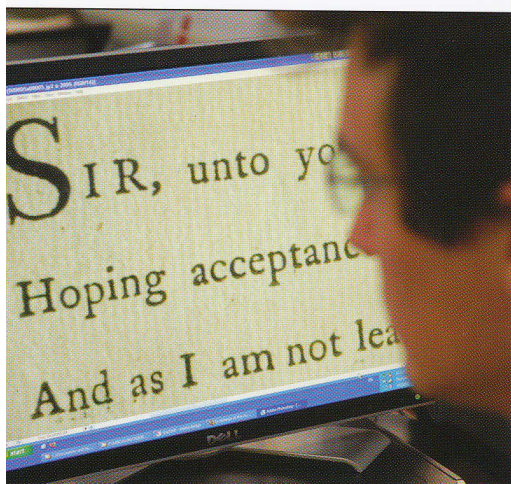


MASTER OF SCIENCE— CONCENTRATION IN DATA CURATION

Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for re-use over time. This new field also includes authentication, archiving, management, preservation, retrieval, and representation.

The Graduate School of Library and Information Science (GSLIS), a top-ranked Library and Information Science (LIS) school, is situated at the University of Illinois, a world-class public research university. The Data Curation (DC) concentration within our ALA-accredited master of science offers a focus on data collection and management, knowledge representation, digital preservation and archiving, data standards, and policy. LIS as a discipline emphasizes the use of information technology to support new approaches to information organization and access. This program provides a strong grounding in the theory and skills necessary to work directly with academic and industry researchers who need data curation expertise.

Students in the DC concentration are taught to plan and manage data curation systems, create and maintain data collections, and to evaluate and apply data and metadata standards for varied uses across the sciences, humanities, and social sciences. Our graduates will be prepared for employment in a range of information-oriented institutions, including museums, data centers, libraries and institutional repositories, archives, and private industry.



COURSEWORK

The master's degree program requires 40 hours of graduate study, including two core courses, "Information Organization and Access" and "Libraries, Information, and Society." Students in the DC concentration will be required to take three additional courses: "Foundations of Data Curation", "Systems Analysis and Management," and "Digital Preservation." Also, participation in an internship or practicum is a cornerstone of the program and is strongly recommended.

For the remaining hours, students in the DC concentration work with their faculty advisors to select electives for a more individualized program that will prepare them for a career focus. A listing of current electives is available at: www.lis.uiuc.edu/programs/ms/data_curation.html.

SCHEDULING OPTIONS

For a variety of reasons, students who want to earn a degree at Illinois are not always able to relocate to campus. We extend our program to them through our online learning option called LEEP. LEEP students begin the program with a 10-day, on-campus summer stay and a brief on-campus stay each semester. Financial support for travel and housing is available for a small number of entering LEEP students each year.

FINANCIAL AID

With the support of the Institute of Museum and Library Services (IMLS), the Data Curation Education Program (DCEP) provides paid internships at our partner organizations' sites. In addition, there will be a limited number of fellowships available for qualified students.

TO APPLY

For an overview of the specific admissions requirements for the M.S., as well as detailed instructions for applying to this program, please see:

www.lis.uiuc.edu/admissions/requirements/ms.html

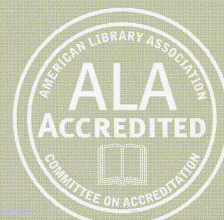
Additionally, incoming students must have basic computer skills, defined at:

www.lis.uiuc.edu/admissions/requirements/tech.html

After reviewing these, prospective students can apply using the University's Web-based electronic application:

www.grad.uiuc.edu/admissions/apply/

For more information, please contact the GSLIS admission office at (800) 982-0914 or lis-apply@uiuc.edu. For specific information about DCEP, please call (217) 244-5574 or browse www.uiuc.edu/goto/dcep





WWW.LIS.UIUC.EDU

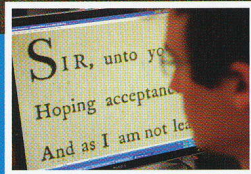
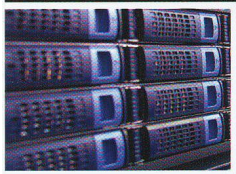
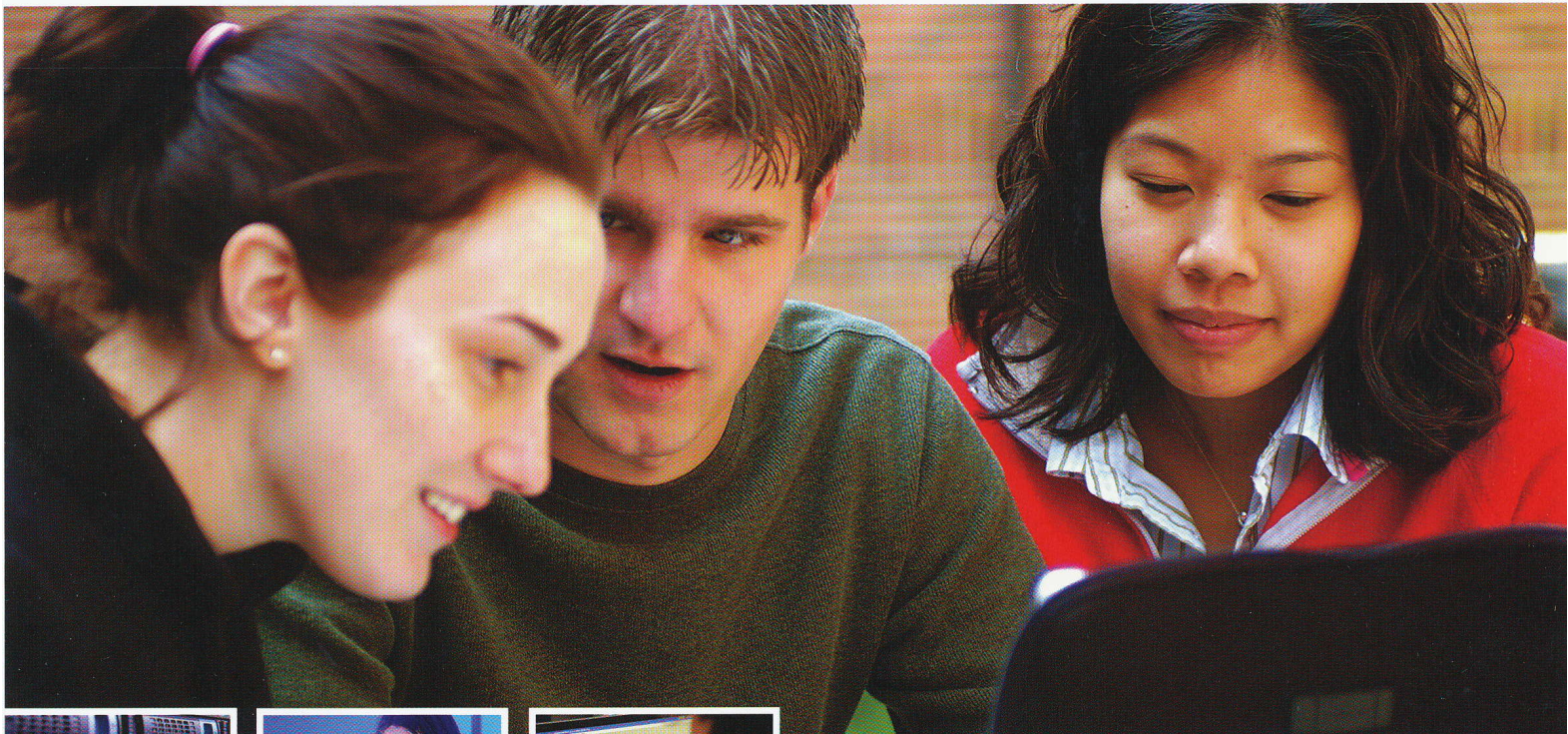
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

501 East Daniel Street
Champaign, Illinois 61820-6211
gslis@uiuc.edu
www.lis.uiuc.edu

Administrative Offices: (217) 333-3280
Admissions: (800) 982-0914
Advancement: (217) 244-9577
Dean's Office: (217) 333-3281
Financial Aid: (800) 982-0914
Fax: (217) 244-3302

GRADUATE SCHOOL OF **LIBRARY AND
INFORMATION SCIENCE**
The iSchool at Illinois

specialization in
data curation



"It is now recognized that data are extremely valuable assets that scientists and scholars can use in new and innovative ways. As more and more research organizations struggle with the flood of digital data, there is growing demand for information professionals to provide curation services. We can't keep all the data, but what we do keep needs to be organized, preserved, and made accessible and usable for the long term."

*Carole Palmer, professor and director
of the Center for Informatics Research
in Science and Scholarship*



UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

specialization in data curation

At its root, the field of library and information science has always been concerned with collecting, preserving, and providing access to information. Every academic field, from the hard sciences to the humanities, produces information in the form of research data that needs to be collected and preserved for future scholars to use. While researchers still generate paper and other analog data formats, it is more common now for data to be digital.

Data curation brings together library and information science and archival theory as well as digital technologies for information discovery, access, and re-use, to maximize the future usefulness of data. Data curators take an active role in the ongoing management of data through its lifecycle of interest to scholarship, science, and education. To do this, they partner with scientists, scholars, computer scientists, IT professionals, and libraries and other organizations that support the research process.

GSLIS has taken a leading role in both data curation education and research, and the program specialization is considered an exemplar in the field. The focus is on data collection and management, knowledge representation, digital preservation and archiving, data standards, and policy.

Fellowships are available to qualified students.

Admission requirements

Admission requirements for the specialization in data curation are the same as those for the master's degree program, which include several steps that allow applicants to show their interest in and qualification for study. Contact the GSLIS Admissions Office with questions about the admissions process at (800) 982-0914 or visit www.lis.illinois.edu/admissions.

GSLIS faculty working in data curation

Melissa Cragin, research assistant professor
(Ph.D. Illinois)

W. John MacMullen, assistant professor
(Ph.D. North Carolina-Chapel Hill)

Jerome McDonough, assistant professor
(Ph.D. Berkeley)

Carole Palmer, professor
(Ph.D. Illinois)

Allen Renear, associate professor
(Ph.D. Brown)

Recent internships

An important element of the specialization is the opportunity for field work. Internships have recently taken place at the following institutions:

- National Snow and Ice Data Center (NSIDC)
- Smithsonian Institution, Digital Services Division
- Purdue University Library, Distributed Data Curation Center
- National Library of Medicine
- State Historical Society of North Dakota
- Brown University Women's Writers Project

CIRSS

The Center for Informatics Research in Science and Scholarship (CIRSS) is the hub of data curation education and research activities at GSLIS. The Center conducts research on information problems that impact scientific and scholarly inquiry. Projects and activities focus on how digital information can advance the work of scientists and scholars, the curation of research data, and the integration of information within and across disciplines and research communities. Find out more at <http://cirss.lis.illinois.edu>

Appendix E: List of Disseminations

Papers (Peer Reviewed)

Collie, W.A., & Witt, M. (2011). A practice & value proposal for doctoral dissertation data curation. Conference paper for the Sixth International Digital Curation Conference, December 6-8, Chicago, IL. Available online at: <http://www.ijdc.net/index.php/ijdc/article/view/189>.

Palmer, C.L., Cragin, M.H., Heidorn, P.B., Smith, L.C. (2007). Data curation for the long tail of science: the case of environmental sciences. Conference Paper at the Third International Digital Curation Conference, December 11-13, Washington, DC. Available online at: https://apps.lis.illinois.edu/wiki/download/attachments/32666/Palmer_DCC2007.pdf.

Heidorn, P.B., Palmer, C.L., Cragin, M.H., & Smith, L.C. (2007). Data curation education and biological information specialists. DigCCurr2007: An International Symposium on Digital Curation, April 18-20, Chapel Hill, NC. Available online at: http://www.ils.unc.edu/digccurr2007/papers/heidornEtal_paper_8-2.pdf.

Palmer, C.L., Heidorn, P.B., Wright, D., & Cragin, M.H. (2006). Graduate curriculum for biological information specialists: a key to integration of scale in biology. *Proceedings of the 2nd International Digital Curation Conference*, November 21-22, Glasgow, Scotland. Available online at: <http://www.ijdc.net/index.php/ijdc/article/view/42>.

Conference Papers and Panels (Peer Reviewed)

Cragin, M.H., Smith, L.C., Palmer, C.L., & Heidorn, P.B. (2009). Extending the data curation curriculum to practicing LIS professionals. Panel for DigCCurr 2009: Digital Curation Practice, Promise and Prospects, April 1-3, Chapel Hill, NC.

McGovern, N.Y., Tibbo, H.R., Cragin, M.H., Davidson, J., & Hofman, H. (2009). Digital curation and preservation training and education: a panel to consider options and intersections. Panel for DigCCurr 2009: Digital Curation Practice, Promise and Prospects, April 1-3, Chapel Hill, NC.

Smith, C.A., Agnew, G., Cragin, M.H., Dryden, J., & Eschenfelder, K.R. (2009). Digital curation policy issues. Panel for DigCCurr 2009: Digital Curation Practice, Promise and Prospects, April 1-3, Chapel Hill, NC.

Palmer, C.L., Cragin, M.H., Tibbo, H., Bradley, J. Botticelli, P. (2009). Emerging education initiatives in data curation. Panel for the Association for Library and Information Science Education Conference, January 22, Denver, CO.

Palmer, C.L. (2008). E-research crosses the pond: contrasting transformations in the U.S. and U.K. Panel for the American Society for Information Science and Technology Annual Meeting (ASIS&T), October 24-29, Columbus, OH.

Palmer, C.L. (2008). Developing curriculum for digital libraries and digital curation education: reflections on synergies and divergencies. Panel for the Society for Information Science and Technology Annual Meeting (ASIS&T), October 24-29, Columbus, OH.

Heidorn, P.B. (2007). Biological information management from molecules to ecosystems. Panel on “Curation Standards and Best Practices,” at 2nd International BioCuration Meeting, October 25-28, San Jose, CA.

Cragin, M.H., D’Avolio, L., MacMullen, W.J., & Smith, C.A. (2007). The effects of context on data quality in biomedical data reuse. Panel at the Annual Meeting of the American Society for Information Science & Technology (ASIS&T), October 19-24, Milwaukee, WI.¹

Heidorn, P.B. & Tibbo, H. (2007). Identifying best practices and skills for workforce development in data curation. Panel at the Annual Meeting of the American Society for Information Science & Technology (ASIS&T), October 19-24, Milwaukee, WI.

Heidorn, P.B., Palmer, C.L., Wright, D., & Cragin, M.H. (2006). Graduate curriculum for biological information specialists: a key to integration of scale in biology. 2nd International Digital Curation Conference, November 21-22, Glasgow, Scotland.
[Revised for publication; see Palmer et al. in Papers section]

Cragin, M.H., MacMullen, W.J., Wallis, J., & Zimmerman, A. (2006). Scholarly communication: scientists’ views of a changing landscape. Panel at the Annual Meeting of the American Society for Information Science and Technology (ASIS&T), November 3-8, Austin, TX.

Posters

Cragin, M.H., Palmer, C.L., Varvel, V.E., Collie, A., & Dolan, M. (2009). Analyzing data curation job descriptions. 5th International Digital Curation Conference, Dec. 2-4, London, UK. Available online at: <http://hdl.handle.net/2142/14544>.

Renear, A.H., Dolan, M., Trainor, K., Cragin, M.H. (2009). Towards a cross-disciplinary notion of data level in data curation. The Annual Meeting of the American Society for Information Science and Technology (ASIS&T), November 8-11, Vancouver, BC. Available online at: <http://www.ideals.illinois.edu/handle/2142/14547>.

¹ Cragin presented “Data and Fitness for Re-use” and MacMullen presented “Gene Ontology Annotations as an Example of the Impact of Curatorial variation on data reuse”.

Palmer, C.L., Renear, A.H., & Cragin, M.H. (2008). Purposeful curation: research and education for a future with working data. 4th International Digital Curation Conference, December 1-3, Edinburgh, Scotland. Available online at: <http://hdl.handle.net/2142/9764>.

Cragin, M.H., Heidorn, P.B., Palmer, C.L., & Smith, L.C. (2007). An educational program on data curation. Presented at the Science and Technology Section program, "Issues and Trends in Digital Repositories of Non-textual Information: Support for Research and Teaching". ALA Annual Conference, June 21-27, Washington D.C. Available online at: <http://hdl.handle.net/2142/3493>.

Palmer, C.L., Cragin, M.H., Heidorn, P.B., Smith, L.C. (2007). Data curation for the long tail of science: the case of environmental sciences. 3rd International Digital Curation Conference, December 11-13, Washington, DC.

Invited Talks and Workshops

Cragin, M.H. & Renear A.H. (2011). Data curation workshop. Johns Hopkins University Libraries, July 7, Baltimore, MD.

Cragin, M.H. (2011). Current developments in data curation for research and learning. Summer In-service program for Visiting Chinese Librarians at the University of Illinois at Urbana-Champaign, June 27, Urbana, IL.

Cragin, M.H. (2011). Growing the curation community in LIS: current research and educational initiatives. Keynote address at the Wisconsin Association of Academic Librarians Annual Conference, April 26-29, Stevens Point, WI.

Palmer, C.L. (2011). Investing in research data. Closing address at the Illinois Research Data Initiative Opening Symposium at the University of Illinois at Urbana-Champaign, September 27, Urbana, IL.

Palmer, C.L., Renear, A.H., McDonough, J., Jewett, B., Smith, A. (2011). Issues and challenges in data management and curation. Panel at the Illinois Research Data Initiative Opening Symposium at the University of Illinois at Urbana-Champaign, September 27, Urbana, IL.

Palmer, C.L. (2010). The data conservancy and data curation research and education at CIRSS. University of Illinois Data Stewardship Committee, November 18, Urbana, IL.

Cragin, M.H. (2010). Data curation: education and current research. School of Information Management, Faculty of Management at Dalhousie University. July 26, Halifax, Nova Scotia.

Cragin, M.H. (2010). Growing the curation community in LIS: data curation education program & the data conservancy. Plenary talk on "Cultural and Organization Initiatives that Meet the Challenges of e-Science" at the 31st Annual IATUL Conference, June 20-24, West Lafayette, IN.

Cragin, M.H. (2009). Data curation in LIS education and libraries. Panel on “Big Science, Little Science, E-Science: The Science Librarian’s Role in the Conversation” for the ACRL Science and Technology Section, July 13, Chicago, IL. Presentation slides available online at: <http://hdl.handle.net/2142/13145>.

Smith, L.C., Cragin, M.H., Heidorn, P. B., Palmer, C.L., MacMullen, W.J. (2008). Developing an effective data curation education program. Workshop on "Education for Digital Stewardship: Librarians, Archivists or Curators?" at the Joint Conference on Digital Libraries (JCDL), June 20, Pittsburgh, PA. Workshop program available at: <http://www.ils.unc.edu/jcdl2008/>.

Palmer, C.L. (2008). Data curation education: evolving research librarianship in the sciences. Librarians & e-Science: Focusing towards 20/20. Committee on Institutional Cooperation, Center for Library Initiatives Conference, May 13, West Lafayette, IN. Video of presentation available at: <http://www-s.cic.net/downloads/conferences/library2008/CarolePalmer.wmv>.

Other Presentations and Outreach:

Cragin, M.H. (2009). Purposeful curation: collections and curation services for data communities. Sheridan Libraries, Johns Hopkins University, August 7, Baltimore, MD.

Cragin, M.H. (2009). Serving scientists and scholars: data practices, collections, and curation. Summer In-service Program for Visiting Chinese Librarians at the University of Illinois at Urbana-Champaign, July 15, Urbana, IL.

Cragin, M.H. (2009). Trends & opportunities: data curation and data literacy. Robert E. Kennedy Library, California Polytechnic State University, June 3, San Luis Obispo, CA.

Cragin, M.H. (2009). Data curation education program: situating data curation at GSLIS. Summer Institute for Data Curation in the Humanities at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign, May 18-22, Champaign, IL.

Cragin, M.H. & Palmer, C.L. (2009). Curation of research data: understanding scholarly practices and collections. Summer Institute for Data Curation in the Humanities at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign, May 18-22, Champaign, IL.

Palmer, C.L. (2008). Data communities: transforming research through “radical sharing.” IT@Illinois Symposium, December 10, Urbana, IL.

Palmer, C.L. (2008). Contouring curation for disciplinary difference and the needs of small science. Sun Preservation & Archiving Special Interest Group meeting, November 19-21, Baltimore, MD. Slides available online at: <http://hdl.handle.net/2142/9715>.

Cragin, M.H. (2008). Data curation services in research libraries. National Library of Medicine, April 15, Bethesda, MD.

Palmer, C.L. (2008). Education for new roles. Presentation at the Association for Research Libraries Fall Forum. October 16-17, Washington, D.C. ARL Fall Forum Proceedings available at: <http://www.arl.org/resources/pubs/fallforumproceedings/forum08proceedings.shtml>.

Service:

Cragin, M.H. & Renear, A.H. [Co-chairs]. (2010). International Digital Curation Conference, December 6-8, Chicago, IL.

Cragin, M.H. [Presenter]. (2010). Introduction to data curation - a running tour through the data curation environment. A pre-conference workshop for the United States Agricultural Information Network (USAIN) Conference, May 9, West Lafayette, IN. Program available online at: <http://usain.lib.purdue.edu/preconf2.html>.

Palmer, C.L., Heidorn, P.B., & Renear, A.H. [Participants]. (2008). Data curation education at Illinois: overview after first year of delivery. International Data Curation Education Action (IDEA) Working Group Meeting, December 4, Edinburgh, Scotland. Conference report available online at: <http://www.dlib.org/dlib/march09/hank/03hank.html>.

Heidorn, P.B. [Presenter]. (2008). Report on the Interagency Working Group on Digital Data: roles of entities, organizations, individuals, roles, sectors, and types. International Data Curation Education Action (IDEA) Working Group Meeting, December 4, Edinburgh, Scotland. Conference report available online at: <http://www.dlib.org/dlib/march09/hank/03hank.html>.

Palmer, C.L. [Participant]. (2007). Sharing and curating research data: meeting programme planning challenges [Workshop], December 14, Washington, D.C.

Palmer, C.L. [Participant]. (2007). Workshop on scientific and scholarly workflow cyberinfrastructure: improving interoperability, sustainability and platform convergence in scientific and scholarly workflow, October 3-5, Baltimore, MD.

Heidorn, P.B. [Co-chair]. (2007). Land informatics symposium and workshops (I and II), September 7-8 and 28-29, Champaign, IL.

Cragin, M.H. & Renear, A.H. [Members]. (2007-2009). Program committee of international digital curation conference.

P. Bryan Heidorn, Ph.D. served as Program Manager of the Division of Biological Infrastructure, National Science Foundation throughout the course of the grant project. He also served on the e-Biosphere Steering Committee and is currently President of the JRS Biodiversity Foundation.

Appendix F: Summer Institute Programs**2008 Summer Institute on Data Curation Schedule**

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
Monday, June 2 8:30 am through Thursday, June 5 3:00 pm

All sessions will be held in **Room 131** at GSLIS unless otherwise noted.

Monday - June 2

- 8:30 - Coffee and Bagels, East Foyer
- 8:45 - Intros and Overview of the week - Melissa Cragin and Carole Palmer (GSLIS)
- 9:30 - Intro to DC and general issues - Melissa Cragin
- 10:15 - Break
- 10:30 - Introduction to Digital Data - W. John MacMullen (GSLIS)
- 12:15 - Lunch - (box lunch at GSLIS)
- 1:30 - Integrity and Authenticity - Allen Renear (GSLIS)
- 3:00 - Break
- 3:15 - Techniques for Evaluating Scientific Data Practices and Needs - Carole Palmer
- 4:30 - Small Group discussion time
- 6:15 - Dinner at the Illini Union

Tuesday - June 3

- 9:00 - Appraisal and Selection - Melissa Cragin
- 10:15 - Break
- 10:30 - Preparing Data for Ingest - Ruth Duerr (NSIDC)
- 12:15 - Lunch (on your own)
- 1:15 - Resource requirements for a data curation program - Scott Brandt (Purdue University Libraries)
- 3:00 - Small Group discussion time
- 4:00 - Librarians and Scientists Working Together - Panel, **Room 126**
- 5:30 - Reception at GSLIS, East Foyer

Wednesday - June 4

- 8:30 - Day-to-Day Digital Preservation - Tim Donohue (UIUC Libraries)
- 10:00 - Break
- 10:15 - Digital Preservation and Standards - Jerome McDonough (GSLIS)
- 12:00 - Lunch (box lunch at GSLIS)
- 1:00 - Digital Preservation and Standards - Hands-On - Jerome McDonough, **Room 52**
- 3:15 - Break
- 3:30 - Technical aspects of DL and Repository systems, architectures and infrastructures
Tim DiLauro (Johns Hopkins Libraries)
- 7:00 - Small Group discussion time

Thursday - June 5

- 8:30 - Free or optional group discussion
- 10:00 - Keynote address - The Island of Our Knowledge, the Shore of Our Ignorance - Anna Gold (Kennedy Library, California Polytechnic State University), Room 126
- 11:30 - Reception, East Foyer
- 1:00 - Group Discussion Outcomes and Wrap-up
- 3:00 - END

2010 Summer Institute on Data Curation

Graduate School of Library and Information Science

University of Illinois at Urbana-Champaign

Monday, May 17 8:30 am through Thursday, May 20 1:30 pm

All sessions will be held in **Room 131** at GSLIS unless otherwise noted.

Monday – May 17: Introductions

- 8:30 - Coffee and Bagels, East Foyer
- 8:45 - Intros and Overview of the week - Melissa Cragin (GSLIS)
- 9:15 - Intro to general Curation Issues for Earth and Environmental Sciences – Ruth Duerr (NSIDC)/ Giri Palanisamy (ORNL) / Bryan Heidorn (UAz)
- 10:45 - Break
- 11:00 - Research and engagement with scientists to advance curation principles and processes - Carole Palmer (CIRSS, GSLIS)
- 12:15 - Lunch - (box lunch at GSLIS)
- 1:30 - What Makes a Data Archive Tick: Marrying Content and User Support – Steve Worley (NCAR RDA)
- 3:00 - Break
- 3:15 – Standards for Earth Science Data – Ruth Duerr
- 5:00 – Wrap-up
- 6:15 - Dinner at the Illini Union, in the General Lounge, 2nd floor North

Tuesday – May 18: Technical Systems

- 8:30 – Coffee and Bagels
- 8:45 – Information Architecture Strategies for Data Curation – John Ferreira (Cornell Univ. Libraries)
- 10:15 – Break
- 10:30 – ORNL Multi-Agency Data Center Projects - A Technical Overview – Giri Palanisamy
- 12:15 – Lunch (on your own)
- 1:30 – An Institutional Perspective on Data Curation: The View from Cornell University Library – Gail Steinhart (Cornell Univ. Libraries)
- 3:00 – Break
- 3:30 – Panel: Library and Data Professionals – Keith Kaneda (JHU Sheridan Libraries), Mary Marlino (Director, NCAR Library), Lynn Yarmey (SCRIPS, LTER), Jake Carlson (Purdue University Libraries). **Held in GSLIS Rm. 126**
- 5:30 – Wrap-up - Dinner on your Own

Wednesday – May 19: GIS data

- 8:30 – Coffee and Bagels
- 9:00 – Making Metadata Work: The NBII Clearinghouse Experience – Vivian Hutchison (USGS / NBII)
- 10:15 – Break
- 10:30 – Q+A with Vivian Hutchison
- 12:00 – Lunch (box lunch at GSLIS)

- 1:00 – Introduction to Geographic Information Systems (GIS) and Light Detection and Ranging (LiDAR) Data Files and Formats –
Sheena Beaverson (Illinois State Geological Survey)
- 3:15 – Break
- 3:30 – LiDAR Data Acquisition, Archiving, and On-line Access – Sheena Beaverson
- 6:30 – “Birds-of-a-Feather” Symposium(s) ALL WELCOME (Lynn Yarmey;
Bryan Heidorn; Melissa Cragin)

Thursday – May 20: Cyberinfrastructure and Curation Research

- 8:45 – Coffee and Bagels
- 9:00 – Comparative Analysis of Environmental Cyberinfrastructures –
Bryan Heidorn
- 10:15 - Break
- 10:30 – GSLIS Research on Data Curation and the Data Conservancy – Melissa Cragin
- 12:00 – 1:30 – Lunch and Wrap-up discussion (box lunch)

2011 SUMMER INSTITUTE SCHEDULE:**Monday – June 6: Introductions**

8:30 - Coffee and Bagels, East Foyer

8:45 - Intros and Overview of the week - Melissa Cragin (GSLIS)

9:00 - Intro to general Curation Issues for Data across the Life Sciences – David "Paddy" Patterson (MBL)
John MacMullen (Illinois); Giovanni Stracquadanio (JHU)

10:45 – Break

11:00 – Joel Bader (JHU)

12:15 - Lunch - (box lunch at GSLIS)

1:30 – Model Organism Data and Community Collections - Nick Stover

3:00 – Break

3:15 – Data Issues in the Life Sciences – Anne Thessen (MBL)

5:00 – Wrap-up

6:15 - Dinner at the Illini Union, in the General Lounge, 2nd floor North

Tuesday – June 7: Consideration of the Technical (Engaging with Scientists and the Technical Side of Data)

8:30 – Coffee and Bagels

8:45 – Research and engagement with scientists to advance curation principles and
processes - Carole Palmer (CIRSS, GSLIS)

10:15 – Break

10:30 – Data set identity issues – Allen Renear (Illinois)

12:15 – Lunch (on your own)

1:30 – The Semantic Web for Life Science Data – Peter DeVries

3:00 – Break

3:30 – Panel: Library and Data Professionals – Lynn Yarmey (Stanford), Leslie Delserone (Univ. of
Nebraska); Jim Kreft (Berkeley, GSLIS MS student); Sarah Wright (Cornell) (Rm. 126)

5:30 – Wrap-up - Dinner on your Own

Wednesday – June 8: Data Use and Users

8:30 – Coffee and Bagels

9:00 – Supporting Data Re-use in Academic Libraries – John MacMullen

10:15 – Break

10:30 – Cross-Community User Requirements and the Biodiversity Heritage Library – Chris Freeland

12:00 – Lunch (box lunch at GSLIS)

1:00 – Data Literacy and Training Undergrads and Grads in Best Practices for Data Management – Carly Strasser

3:15 – Break

3:30 – Data Management Planning – Leslie Delserone & Melissa Cragin

6:00ish – gather at dinner location(s), and “Birds-of-a-Feather” Symposium(s) ALL WELCOME (Topic Leaders: Lynn Yarmey; tbd; Melissa Cragin)

Thursday – June 9: Emerging Infrastructures

8:45 – Coffee and Bagels

9:00 – DataNet - Emerging Curation Infrastructure – Melissa Cragin & Carly Strasser (National Center for Ecological Analysis and Synthesis)

10:15 - Break

10:30 – Report on Birds of a Feather discussions; action planning for return to home institutions

12:00 – 1:30 – Lunch and Wrap-up discussion (box lunch)